NAVSHIPS 0967-129-3010

INTRODUCTION TO

SONAR TECHNOLOGY

DECEMBER, 1965

BUREAU OF SHIPS-NAVY DEPARTMENT-WASHINGTON,D.C.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C., 20402 - Price $2.25

PREFACE

The effort expended in the area of Sonar Technology has grown considerably during the past twenty-five years. However, documents covering the knowledge accumulated on sonar technology have not kept up with the work done.

This document, prepared by TRACOR, INC. under the direction of Dr. A. F. Wittenborn, is the outgrowth of an indoctrination program for new employees who were to begin working in the area of anti-submarine warfare at the Bureau of Ships. While not intended as a basic reference document for the design of sonar systems, it is hoped that this publication will serve to highlight the significant areas with which the worker must be concerned if he is to participate in the further advancement of sonar technology.

Washington, D.C.

February, 1966

S.W.W. SHOR, CAPT, USN
Director, Sonar Systems Office
Bureau of Ships

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

TABLE OF CONTENTS (Cont'd)

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF ILLUSTRATIONS (Cont'd)

LIST OF ILLUSTRATIONS (Cont'd)

1. INTRODUCTION TO SONAR SYSTEMS

1.1 DETECTION OF UNDERWATER TARGETS

A number of methods of detecting underwater targets have been employed, depending upon the tactical situation. The most successful have relied upon the following:

1. Magnetic signatures
    a. Self-generated
    b. Perturbation of earth's magnetic field
2. Optical signatures
3. Electric field signatures
4. Thermal signatures (infra red)
5. Hydrodynamical (pressure) signatures
6. Acoustic signatures
    a. Radiated self-noise
    b. Reflected energy

Of these detection methods, the acoustic signatures have been exploited in military sonar.

1.2 FUNDAMENTAL DEFINITIONS

Sonar, an acronym for SOund NAvigation and Ranging, designates that branch of applied acoustics in which acoustic energy is propagated through a water medium. Systems which utilize underwater acoustic energy for observation or communications are referred to as sonar systems.

Sonar systems, as defined above, are used for many purposes ranging from peaceful "fish finders" and small boat navigation aids to large antisubmarine warfare (ASW) systems for detection and classification of submarines, and mine hunting. Sonar systems also provide a means for both short and long distance underwater communications.

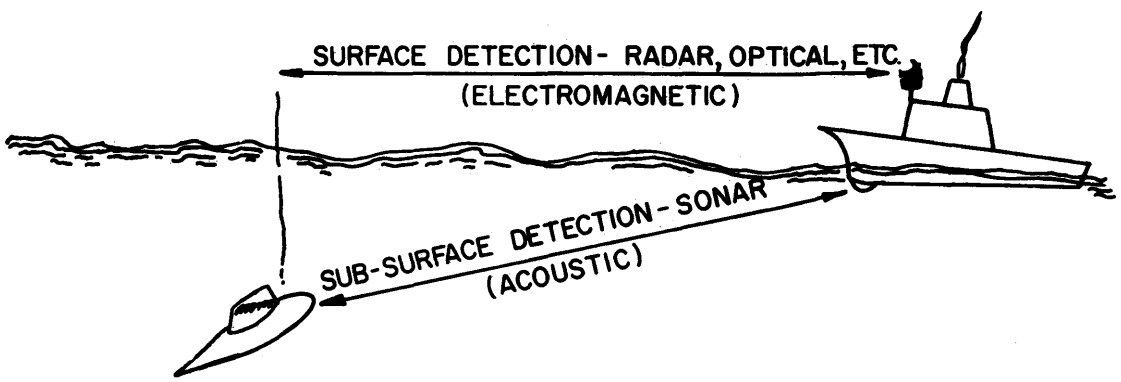
Why choose sonar to perform these jobs when radio and radar systems seem to work so well? Primarily because of the extremely high attenuation and scattering that electromagnetic energy suffers in a highly conductive medium such as sea water. A simple comparison of electromagnetic and acoustic waves is shown in Fig. 1-1. Electromagnetic energy has a velocity some 200,000 times greater than underwater sound; it is attenuated in water at a much greater rate than acoustic energy, namely about $1.3 \times 10^3 f^{1/2}$ dB for each one thousand yards of transmission while acoustic energy is attenuated at about $0.01 f^2$ dB for each one thousand yards, where f is expressed in kcs. The severe attenuation of electromagnetic waves in the conductive oceans makes an underwater "radar" no match for an acoustic system.

1.3 REFERENCES

Before continuing in the discussion of sonar systems and underwater sound, it should be pointed out that the discussion contained in these sections is necessarily limited in scope. A number of references are available which cover the entire field in considerable detail. The following is a list of references which will be found helpful for those who desire a more detailed and comprehensive survey of underwater acoustics and sonar.

1. J. W. Horton, Fundamentals of Sonar, Annapolis, Md., U. S. Naval Institute, 2nd Ed., 1959.
2. C. B. Officer, Introduction to the Theory of Sound Transmission, McGraw-Hill, 1958.
3. Vernon M. Albers, Underwater Acoustics Handbook, The Pennsylvania State University Press, 1960.
4. National Defense Research Committee, Physics of Sound in the Sea, (NDRC Div. 6 Summary Tech. Rept. Vol. 8), 1946.

C_E = 300,000,000 METERS/SEC



C_S = 1500 METERS/SEC

$L_S = 0.01 f^2$ DB/KILOYARD

$L_E = 1.3 \times 10^3 f^{1/2}$ DB/KYD
(IN WATER)

$$\frac{C_E}{C_S} = 200,000 \quad \text{(VELOCITY RATIO)}$$

$$\frac{L_S}{L_E} = 7 \times 10^{-6} f^{3/2} \text{ (RATIO OF ATTENUATION COEFFICIENTS)}$$

FIG. I-1 -COMPARISON OF ELECTROMAGNETIC AND ACOUSTIC PROPERTIES

5. Philip M. Morse, Vibration and Sound, McGraw-Hill, Second Ed., 1948.

6. Leo L. Beranek, Acoustic Measurements, John Wiley & Sons, 1949.

7. R. J. Urick and A. W. Pryce, A Summary of Underwater Acoustic Data, Part I-VIII, Office of Naval Research, 1953-1956. (Classified)

8. National Defense Research Committee, Principles and Applications of Underwater Sound, NDRC Div. 6, Summary Tech. Rept. Vol 7, 1946.

1.4 THE SONAR EQUATION

The fundamental problem in any detection system is to find some way of separating information indicating the presence of the object to be detected from the background or clutter caused by similar pieces of information completely unrelated to the object to be detected. Specifically, a sonar system attempts to do this using acoustic energy (signal) either radiated by or reflected from a target. In general, there are other sources of acoustic energy in the ocean which tend to mask the signal. A great deal depends upon a quantity known as signal-to-noise ratio $[\frac{S}{N}]$[1]. In this ratio [S] is the average signal power in watts measured over the time it exists and [N] is called the average noise power in watts in the band of the sonar (It refers to the average power in the interfering, background waveform.). The signal and noise powers are the per unit area values in the water just outside the transducer. It is generally assumed that a signal may be detected on the average 50% of the time in the presence of noise if the signal-to-noise ratio exceeds a number [M] known as the recognition differential. We write formally then

[1] Square brackets are used in this section to indicate power.

$$\left[\frac{S}{N}\right] \geq [M] \quad , \tag{1-1}$$

if the probability that a detection will be made is greater than or equal to 0.5. This assumption is a convenient approximation useful in discussing detection systems; [M] is dependent upon the signal processing, the display, and the decision rule employed in detection. The role of [M] will be discussed further in Chapters 2 and 6 entitled "Sonar Parameters" and "Signal Processing."

As defined, the signal-to-noise ratio $\left[\frac{S}{N}\right]$ and [M] represent the ratios of two powers. It is customary in describing sonar system behavior to express power ratios in decibels (dB). The decibel is a logarithmic unit which is defined as 10 times the logarithm to the base 10 of the ratio of two powers. Suppose 1 watt is taken as a reference power to which all others are to be compared, then the signal power S expressed in decibels would be[2]

$$S = 10 \log \frac{[S] \text{ watts}}{[1] \text{ watt}} \quad . \tag{1-2}$$

In the same way

$$N = 10 \log \frac{[N] \text{ watts}}{[1] \text{ watt}} \quad , \tag{1-3}$$

and

$$M = 10 \log [M] \quad . \tag{1-4}$$

The relationship between S, N, and M which is consistent with (1) is

$$S - N \geq M \tag{1-5}$$

---

[2]Throughout the book, the symbol, log, is used to denote logarithms to base 10.

because

$$10 \log \frac{[S] \text{ watts}}{[1] \text{ watt}} - 10 \log \frac{[N] \text{ watts}}{[1] \text{ watt}} \geq 10 \log [M]$$

is equivalent to Eq. (1-5) and the sum property of logarithms provides the simplification to the form,

$$10 \log \left[\frac{S}{N}\right] \geq 10 \log [M] \quad ,$$

which is equivalent to Eq. (1-1).

Equation (1-5) is the basis for setting up all of the versions of the sonar range equation used in comparing sonar performance. The requirement to be met is that the signal, expressed in decibels, minus the noise expressed in decibels, must be equal to or greater than a number M called the recognition differential, also expressed in decibels, if the detection probability is to be 0.5 or greater. It is repeated here for emphasis:

$$S - N \geq M \quad . \qquad (1\text{-}6)$$

This expression is the basis for all forms of the sonar equation. It must be remembered that M is not a constant; in general, its value will depend upon the number of clutter points (false target points) on the display which can be dealt with.

1.5 THE LISTENING EQUATION FOR PASSIVE SYSTEMS

A passive sonar depends on receiving a signal which is radiated by the target. The target signal can be caused by operating machinery, propeller noise, hull flow noise, etc., but

the same fundamental signal-to-noise ratio requirement must be satisfied. The passive equation starts with

$$S - N = M \quad , \tag{1-7}$$

in which S is the signal level and N is the noise level in the water just outside the transducer.[3] If it is assumed that the target radiates an acoustic signal at a level of T_p dB, it will arrive at the receiving array of the detecting ship reduced in level by H dB, the loss experienced in traveling through the water. The signal level, S, at the detecting ship, is given by

$$S = T_p - H \quad . \tag{1-8}$$

Noise, N, acts as a masking signal and is not wanted. Therefore, we design a transducer, composed of many hydrophones, sensitive primarily in the direction of the target so that we can discriminate against the noise as much as possible. This discrimination against noise can be referred to as a spatial processing gain obtained by making the beam pattern (the transducer spatial response) large primarily in the direction of the signal source. This gain is expressed as the receiving directivity index Δ_R. Δ_R gives the reduction in noise level, in dB, obtained by the directional properties of the transducer array. In our basic equation therefore, noise is now altered by the term Δ_R so that the noise term becomes

$$N + \Delta_R \quad . \tag{1-9}$$

Δ_R is always a negative quantity so that $N + \Delta_R$ is always less than N.

---

[3] The level of a quantity in underwater acoustics is 10 times the logarithm of that quantity to a reference (0 dB) quantity. Usually, the 0 dB reference is taken as the power or intensity associated with a sound pressure of 1 dyne/cm^2.

The passive sonar equation can now be constructed in terms of signal and noise. When S and N are substituted from Eqs. (1-8), (1-9) into (1-7) the result is

$$T_p - H - N - \Delta_R \geq M \quad . \tag{1-10}$$

Equation (1-10) is the simplest form of the passive sonar equation. It is useful for estimating the range at which a target can be detected.

T_p is called the target strength

H is the propagation loss

Δ_R is the receiving directivity index.

Inequality (1-10) is usually written as an equality, it being understood that when the left member exceeds M detection is highly probable and when it is less than M, detection seldom occurs. In order to estimate detection range, the functional form of H and values of T_p, N, M, and Δ_R must be obtained.

1.6 THE ECHO-RANGING OR ACTIVE SONAR EQUATION

In an active sonar, acoustic energy is transmitted and the signal is an echo from the target. Two equations are needed to describe the active sonar: one for an isotropic noise limited situation and the other for the reverberation limited situation. Again, the requirement that the signal-to-noise ratio must be greater than or equal to the recognition differential governs the sonar performance. The difference in the two equations depends upon the characteristics of noise and reverberation; noise is to a good approximation describable as isotropic (as much noise power arrives from one direction as any other on the average) while reverberation returns primarily from the direction in which the sonar has transmitted or pinged.

If a sonar transmits an acoustic pulse with an initial source level of S_o dB, the transmitted pulse will suffer a transmission loss of H dB in traveling to the target. The target will scatter acoustic energy, some of which will return to the sonar. The back scattered intensity is related to the target strength, T dB (T itself is related to the so-called scattering cross section of the target). The returning echo will again undergo a propagation loss of H dB. The echo level (or signal level) at the sonar will be given by

$$S_o + T - 2H \quad , \tag{1-11}$$

since the transmission loss in this case is twice the one-way transmission loss.

Under some conditions the background due to the initial transmission (reverberation) will have disappeared and only the ambient noise will be present. The noise or masking level will be identical to that described in the passive listening case and the fundamental signal-to-noise ratio can then be expressed as

$$(S_o + T - 2H) - (N + \Delta_R) = M_n \quad . \tag{1-12}$$

Equation (1-12) is the basic echo ranging equation which is used when the sonar is operating in a noise limited situation. Specific expressions can be inserted in (1-12). Details of these expressions will be found in Chapter 2.

If on the other hand, the echo returns when the reverberation background has not decayed to a level below the ambient noise, the background level is given by Q, the reverberation level at the assumed echo arrival time and the sonar equation takes the form,

$$(S_o + T - 2H) - Q = M_r \quad , \qquad (1\text{-}13)$$

where M_r is the recognition differential for reverberation. The Δ_R term disappears in the reverberation limited case, because reverberation is non-isotropic. All terms in these equations are in dB. Equation (1-13) is known as the reverberation limited echo ranging equation. This expression can be written in terms of the specific functions for S_o, T, and H. Inclusion of an expression for Q is difficult. Approximate models for the time variation of Q will be found in Chapter 12 of Reference 4.

1.7 FIGURE OF MERIT

The figure of merit (FOM) of a sonar system is the maximum transmission loss which it can tolerate and still provide the necessary recognition differential. From Eq. (1-12) for the active sonar,

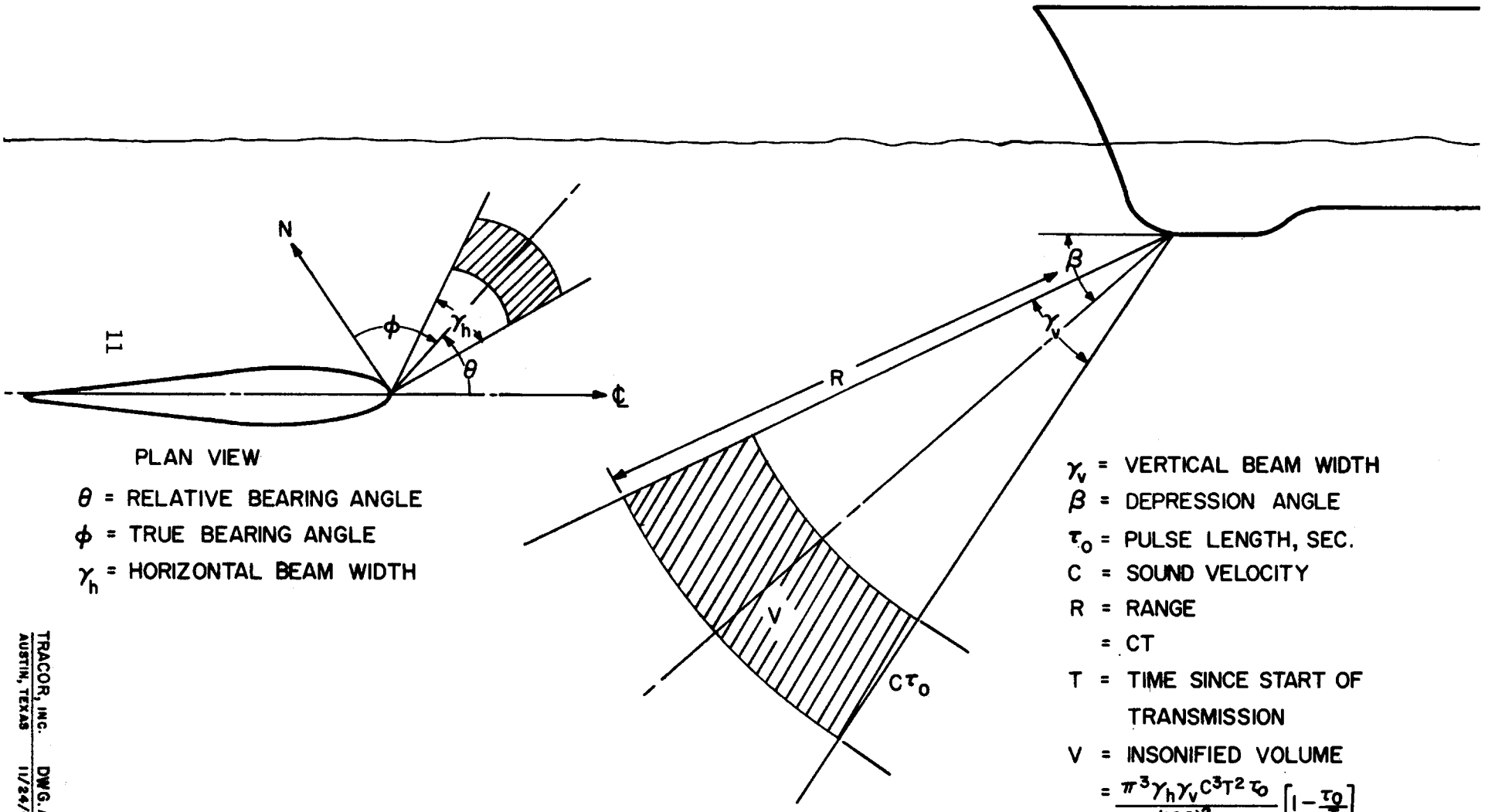
$$FOM = |2H - T|_{max}$$

$$= S_o - (N + \Delta_R) - M_n \quad . \qquad (1\text{-}14)$$

FOM is improved by raising source level S_o (this can be accomplished by increasing the transmitted power level), decreasing the ambient noise level, increasing the absolute value of the receiving directivity index, and decreasing the recognition differential. FOM is used to compare sonars or estimate maximum detection range.

1.8 SONAR GEOMETRY

The geometry of the active sonar operation is described in Fig. 1-2. For this discussion the beam has been assumed to have vertical width γ_v, horizontal width γ_h and negligible side



θ = RELATIVE BEARING ANGLE

ϕ = TRUE BEARING ANGLE

γ_h = HORIZONTAL BEAM WIDTH

γ_v = VERTICAL BEAM WIDTH

β = DEPRESSION ANGLE

τ_0 = PULSE LENGTH, SEC.

C = SOUND VELOCITY

R = RANGE

= CT

T = TIME SINCE START OF TRANSMISSION

V = INSONIFIED VOLUME

$$= \frac{\pi^3 \gamma_h \gamma_v C^3 T^2 \tau_0}{(180)^2} \left[1 - \frac{\tau_0}{T}\right]$$

FIG.I-2 - SONAR GEOMETRY

TRACOR, INC. AUSTIN, TEXAS DWG. A769-54 11/24/64 JC/LBH

lobe structure. The beam widths are usually taken as the angles between 6 dB down points. In the most general case, the axis of the beam is depressed with respect to the horizontal by an angle β in a vertical plane rotated θ relative to the axis of the ship. Ship heading and θ determine the true bearing of the axis of the sonar beam. If τ_o is the duration of transmission (the pulse length) the volume of insonified water is

$$V = \frac{\pi^3 \gamma_v \gamma_h}{(180)^2} c^3 T^2 \tau_o \left(1 - \frac{\tau_o}{T}\right) .$$

In this expression T is the elapsed time since the transmission began. The extent in range of the insonified region is $c\tau_o$. The maximum range separation of two scatterers which can simultaneously produce return to the sonar is known as the ping length:

$$\text{ping length} = \frac{c\tau_o}{2} .$$

When a transmission is made, the acoustic energy propagates by one or more modes to other parts of the ocean.

1.9 THE SURFACE CHANNEL MODE

The presence of an isothermal layer near the surface results in refraction paths which tend to bend the acoustic rays so that they are concave upward. The waves strike the surface and are reflected as indicated in Fig. 1-3. Only those rays which strike the surface at a small grazing angle can remain in the surface channel; those with too large a grazing angle strike the lower surface of the isothermal layer and are ~~most~~ lost from the surface channel. Scattering by inhomogeneities in the volume of the channel and at the surface also contributes to the leakage from the channel. This mode of propagation is known as refracted-surface reflected (RSR) propagation.

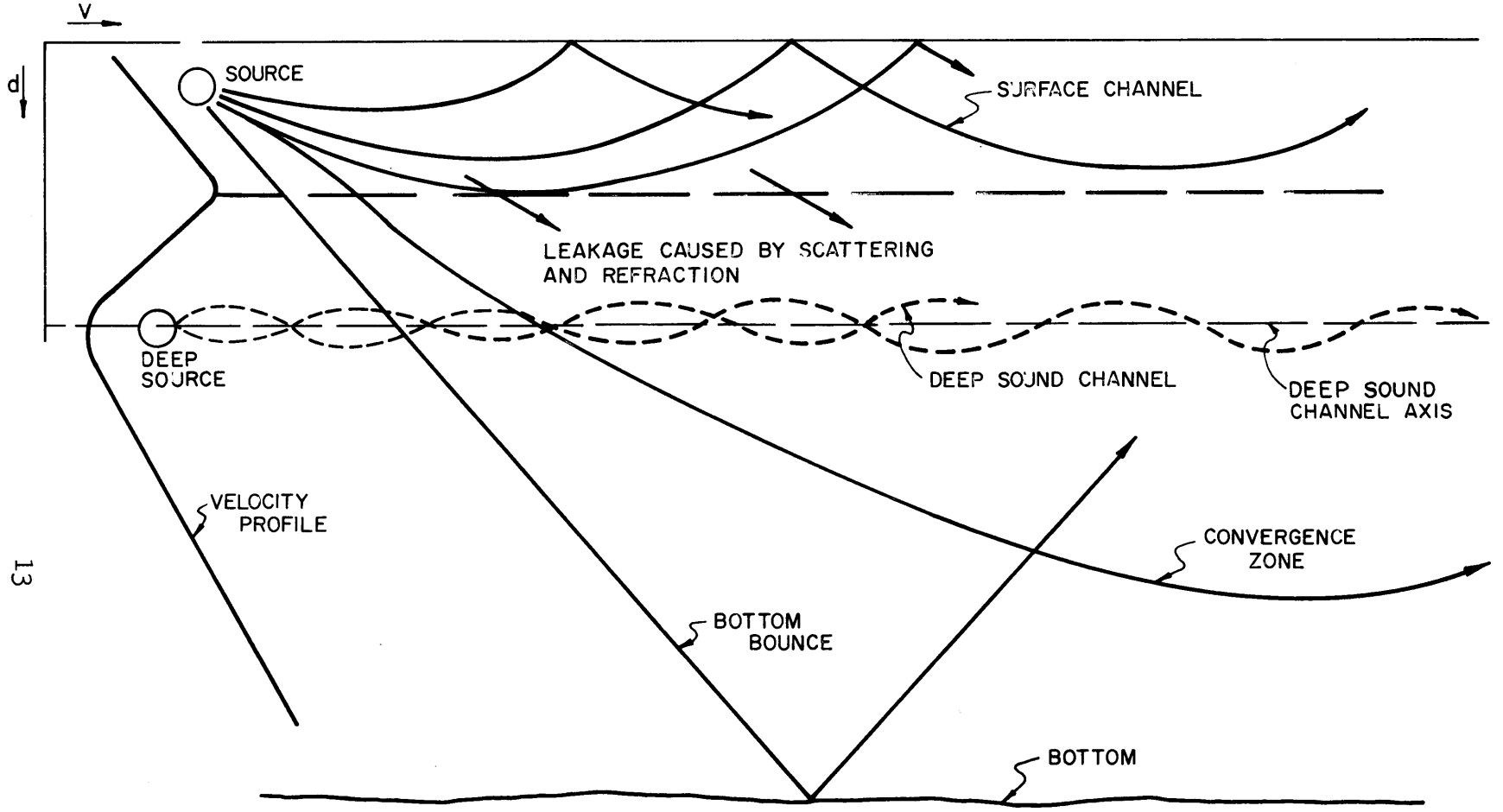


FIG. I-3- MODES OF SOUND TRANSMISSION

TRACOR, INC. | DWG. 771-23
AUSTIN, TEXAS | 5-25-65 BMB-ES

1.10 DEEP PROPAGATION MODES

Some acoustic energy finds its way to the region below the isothermal layer. When surface mode propagation is desired, this is a disadvantage because acoustic energy is lost. There are deep propagation modes which are important to modern sonar. In attempting to use them, acoustic energy is directed downward. Beam depressions from 5^{o} down to 45^{o} are normally employed.

The bottom bounce mode is the simplest to describe. Acoustic energy is directed at the bottom, is reflected there toward a volume near the surface where target submarines may be found. Scattered waves from the target return to the sonar via the same route. Passive detection of the target is also practical using the bottom bounce mode. By employing a maximum depression angle β of 45^{o}, the shortest horizontal range R_h to the target is about

$$R_h \doteq \frac{2D}{\text{Tan } \beta} = 2D$$

where D is the water depth. For average water depths, the shortest bottom bounce operational horizontal range approximately matches the normal longest operational range of the surface mode. Decreasing the depression angle increases the range out to four or fives times the short range limit.

When the depression angle is decreased below a critical value, refraction in the deep water turns the acoustic rays concave upward enough that reflection from the bottom does not occur. This path, denoted in Fig. 1-3 by "convergence zone," eventually returns to the surface at a distance somewhat greater than the longest bottom bounce range. Convergence zone (CZ) propagation requires deep water. The minimum depth required is determined by the local acoustic velocity profile.

The deep sound channel occurs at the minimum in the velocity profile. Acoustic energy is trapped in this channel by the refraction properties of the velocity minimum. Propagation of acoustic energy from explosive sources has been observed over thousands of miles in this channel. To obtain significant propagation in the deep sound channel, the acoustic source must be located within the channel.

1.11 SONAR TERMINOLOGY

Terms and units which have been introduced so far, and which will be used in subsequent chapters, are explained below.

1.11.1 Sound Velocity

Equivalent values for the velocity of sound in sea water near the surface (temperature - 60°F; salinity 34°/oo[4]) are

$$
\begin{aligned}
C &= 4935 \text{ feet/second} \\
&= 1645 \text{ yards/second} \\
&= 1504 \text{ meters/second} \\
&= 1.504 \times 10^5 \text{ centimeters/second} \\
&= 2960 \text{ knots} \\
&= 3400 \text{ mph} \quad .
\end{aligned}
$$

1.11.2 Acoustic Pressure

The acoustic pressure is the deviation from hydrostatic pressure caused by the presence of sound. Pressure is measured in dynes/cm^2 ≡ μbar.

1.11.3 Specific Acoustic Impedance

The specific acoustic impedance is the ratio of acoustic pressure p and medium particle velocity u. To a good approximation

$$\frac{p}{u} = \rho_o c \quad ,$$

[4] o/oo ≡ parts per thousand

where ρ_o is the density of sea water and c is the velocity of sound in sea water. The units of ρ_o c will be

$$\left(\frac{g}{cm^3}\right)\left(\frac{cm}{sec}\right) = \frac{g}{cm^2\ sec} \equiv \text{acoustic ohms} .$$

The units of p/u will be the same:

$$\left(\frac{dyne}{cm^2}\right) \Big/ \left(\frac{cm}{sec}\right) = \left(\frac{g\ cm}{sec^2\ cm^2}\right) \Big/ \left(\frac{cm}{sec}\right) = \frac{g}{cm^2\ sec}$$

1.11.4 Acoustic Intensity

The average rate at which energy crosses unit area is called the acoustic intensity. The instantaneous energy rate is pu: Since $u = p/\rho_o c$, the instantaneous energy rate can be written $p^2/\rho_o c$; the intensity I is

$$I = p^2/\rho_o\ c \quad .$$

An equivalent expression is

$$I = \rho_o\ c\ u^2 \quad .$$

1.11.5 Transducer Output

The rate P (power) at which energy emerges at the electrical output of a transducer array is proportional to the rate at which acoustic energy arrives at the array. The power P is related to the average square of the waveform w(t) (potential difference or current) at the transducer output.

$$P = k\ I \quad ,$$

$$P = K \overline{w^2(t)} \quad .$$

K and k are two constants which can be determined if necessary. K depends upon the transducer impedance and the load impedance; k depends upon the transducer aperture, the beam pattern, and the angle between the beam axis and the direction to the acoustic source. I is the acoustic intensity just outside the transducer.

1.11.6 The Decibel

The ratio of two average powers is often expressed in decibels. If $[P_1]$ and $[P_o]$ are two average powers, their ratio in decibels is

$$10 \log \frac{[P_1]}{[P_o]} \quad .$$

(Only logarithms to base 10 are used in dB representation.) If P_o is the agreed standard power to which all other powers are to be compared, then $[P_1]$ expressed in decibels (denoted by P_1 without brackets) is

$$P_1 = 10 \log \frac{[P_1]}{[P_o]} \quad .$$

P_1 is the dB level of power $[P_1]$ referred to $[P_o]$. $[P_o]$ has been chosen as 1 milliwatt in some disciplines, 1 watt in others. The reference $[P_o]$ usually employed in underwater acoustic intensities is the intensity corresponding to an acoustic wave whose root-mean-square pressure[5] is 1 dyne/cm^2. Intensities in the water can also be expressed in terms of the dB level.

[5] In some older publications, the airborne acoustic reference, 0.0002 dynes/cm^2, is used.

$$I_1 = 10 \log \frac{[I_1 A]}{[I_o A]} = 10 \log \frac{[I_1]}{[I_o]} \quad .$$

It is often desirable to determine acoustic intensity level directly from the root-mean-square (rms) pressures p_1 and p_o

$$I_1 = 10 \log \frac{[Ap_1^2/\rho_o c]}{[Ap_o^2/\rho_o c]} = 10 \log \frac{[p_1^2]}{[p_o^2]} \quad .$$

The logarithm of the square of a number is twice the logarithm of the number so that

$$I_1 = 20 \log \frac{[p_1]}{[p_o]} = 20 \log [p_1] \quad .$$

In this expression p_o, the reference pressure, has been put equal to the standard value, 1 dyne/cm^2.

The most convenient dB computer is a slide rule with an L (logarithm to the base 10) scale and a D scale. One is shown in Fig. 1-4. To convert a number $(N_1) = 6{,}900{,}000$ to dB level, it is first written in the form

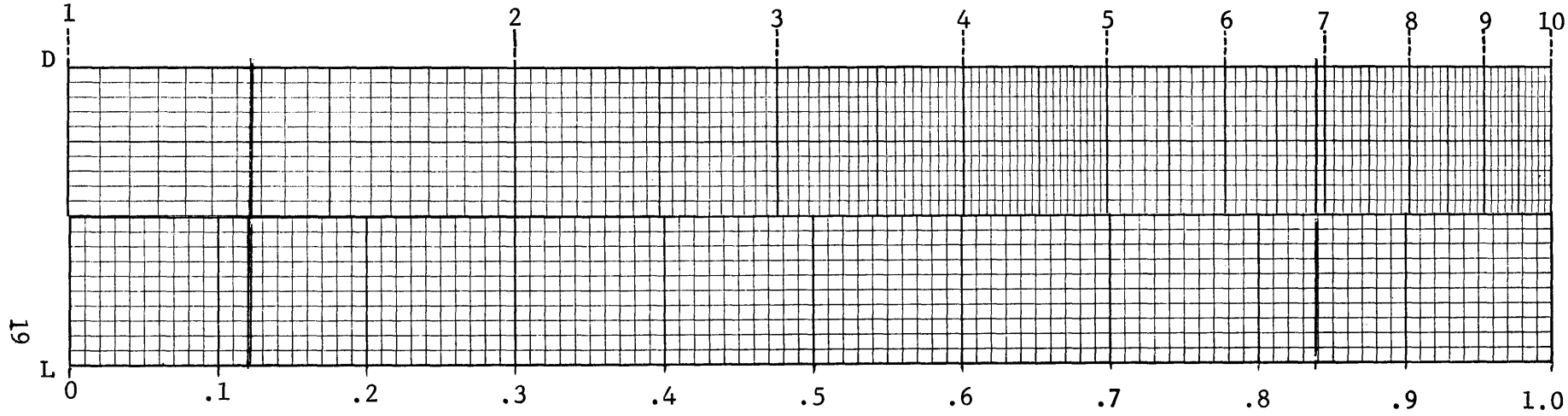
$$6.9 \times 10^6 \quad .$$

The quantity N_1(dB) is then

$$N_1(\text{dB}) = 10 \log (6.9 \times 10^6)$$

$$= 10 \log 6.9 + 10(6) \quad .$$

[N], POWER, INTENSITY, (PRESSURE, VOLTAGE)



LOGARITHM

Write $[N_1]$ in the form 6.9×10^6

$[N_2]$ in the form 1.32×10^{-4}

$N_1 = 10 \log [N_1] = 10 \log (6.9 \times 10^6) = 10 \log (6.9) + 10 \log (10^6) = 10(.84) + 10\ (6) = 68.4$ dB

$N_2 = 10 \log [N_2] = 10 \log (1.32 \times 10^{-4}) = 10 \log (1.32) + 10 \log (10^{-4}) = 10(.121) + 10(-4) = -38.79$dB

When $[N_1]$ is a pressure or voltage the multiplier 10 is replaced by 20.

Fig. 1-4. Conversion to dB Level

The term log 6.9 is obtained by entering the D scale (Fig. 1-4) with 6.9 to obtain 0.84. Log $10^6 = 6$ so that

$$N_1(\text{dB}) = 10\ (0.84) + 10(6) = 68.4\ \text{dB}.$$

A second example, one showing how to obtain the dB level of a fractional ratio, is also shown in Fig. 1-4. It is useful to remember that

a factor 2 in power is +3 dB
a factor 0.5 in power is -3 dB
a factor 10 in power is +10 dB
a factor 0.1 in power is -10 dB
a factor 2 in pressure is +6 dB
a factor $\sqrt{2}$ in pressure is +3 dB .

These and other factors of interest may be read from Fig. 1-4.

2. THE SONAR PARAMETERS

2.1 THE SONAR RANGE EQUATION

The sonar range equation provides an estimate of the maximum range at which a target can be detected for specified recognition differential, target strength, and noise level.

The general equation for the active sonar, noise limited case is

$$2[20 \log R + aR] + \varphi = S_o + T - \underbrace{n - 10 \log w - \Lambda_R}_{\text{total noise term}} - M_n , \qquad (2\text{-}1)$$

where

R = Range in yards
a = Attenuation - dB/yd
φ = Refraction loss - dB
S_o = Source level - dB re one yard
T = Target strength - dB
n = Spectrum noise level - dB
w = Effective bandwidth - cps
Λ_R = Receiving directivity index - dB
M_n = Recognition differential for noise - dB, at the beamformer output

This equation is identical to Eq.(1-12) except for the substitution of a specific set of parameters. The individual parameters are discussed in detail in this section.

For the reverberation limited case

$$2[20 \log R + aR] + \varphi = S_o + T - Q - M_R , \qquad (2\text{-}2)$$

where

Q = Reverberation level - dB

M_R = Recognition differential for reverberation - dB, at the beamformer output

2.2 RANGE, R

The range to a target from the sonar transducer given by either of these equations is the maximum slant or "best straight line" range at which the recognition differential can be obtained. It is a function of the surface range, sonar depth, and target depth. Range is usually measured in yards or kyds. The "range" indicated on a sonar range display is not necessarily the actual range to the target. Acoustic energy may travel by a variety of paths in reaching a target and returning. Refraction and reflection must be considered. In bottom bounce operation, the actual travel time (and therefore measured range) may be considerably greater than the actual slant range. The range required in practice is the horizontal range. The propagation geometry must be known in order to compute an accurate value of horizontal range.

2.3 SPREADING LOSS, 20 log R

From purely geometrical considerations, we know that the decrease in intensity of an acoustic wave traveling a distance R in the medium varies inversely as R^2. If sound emanates from a point, the acoustic wave front at any instant is a spherical surface of radius R from the point source. At a distance R, the surface area of the sphere is $A = 4\pi R^2$ and, of course, A increases as R increases. However, the total power through each spherical surface remains constant, where

Power = Intensity x Area

At R_o = 1 yd, let I_o be the acoustic intensity; then if at R, the intensity is I

$$I_o(4\pi R_o^2) = I(4\pi R^2) \qquad (2\text{-}3)$$

It follows that I varies inversely as R^2:

$$I = I_o/R^2 \qquad (2\text{-}4)$$

When Eq. (2-4) is converted to decibels, the result is

$$I(dB) = I_o(dB) - 20 \log R \qquad (2\text{-}5)$$

The term, -20 log R, is usually referred to as "spreading loss". This is not a loss in the usual sense of the word; it merely represents the decrease in power arriving per unit area as a function of range.

If the acoustic wave is propagated in a confined channel, such as the surface duct, the wave front spreads cylindrically rather than spherically and the relation becomes

$$I(dB) = I_o(dB) - 10 \log R \quad . \qquad (2\text{-}6)$$

For use in paragraph 2.4, it will be useful to consider the derivative of I with respect to R. From Eq. (2-4),

$$\frac{dI}{dR} = -\frac{2I_o}{R^3} = -\frac{2I}{R} \qquad (2\text{-}7)$$

2.4 ATTENUATION COEFFICIENT, a

Attenuation results from any mechanism which removes energy from an acoustic wave. The so-called "spreading loss" is not a loss in the sense that energy is removed from the beam. The energy radiated per unit solid angle would be constant were there no "losses". Losses are caused by scattering from the beam and "heating".

In general, the rate of change in intensity I with range is approximately

$$\frac{dI}{dR} = - \alpha I - \frac{2}{R} I \ . \qquad (2\text{-}8)$$

The second term on the right in Eq. (2-8) is identical to the right member of Eq. (2-7). When attenuation occurs, an additional term $-\alpha I$, which causes I to decrease with R must be added.

Integration of this equation gives

$$I = \frac{Ae^{-\alpha R}}{R^2} \ , \qquad (2\text{-}9)$$

where A is an arbitrary constant to be fixed by the source strength. Ten times the logarithm of both sides of Eq. (2-9) gives

$$\begin{aligned} 10 \log \frac{I}{A} &= - 10[\log R^2 - \alpha R \log e] \\ &= - 10 \log R^2 - aR \\ &= - 20 \log R - aR \ , \end{aligned} \qquad (2\text{-}10)$$

where $a = 10\alpha \log e$ has been written. The quantity a is called

the attenuation coefficient. When $\alpha << 1/yd$, then A is almost equal to I_o, the intensity at R = 1 yd. The first term in Eq. (2-10) is called the spreading loss; the second, the attenuation loss.

From empirical studies, the attenuation coefficient is satisfactorily described by the expression,

$$\alpha = \frac{40\ f^2}{4100 + f^2} + 0.000275\ f^2\ \text{dB/kyd}, \qquad (2\text{-}11)$$

where f is in kcs. At low frequencies, this reduces to the $0.01\ f^2$ value given in paragraph 1.2. A number of different investigators have measured attenuation in the past, and several different expressions for the attenuation coefficient can be found in the literature. The largest variations in values obtained from these different expressions occur at the low frequency end of the spectrum. In a recent paper, Thorp[1] has reported that for frequencies of 112 cps to 1780 cps, the attenuation coefficient is given by

$$a = 5.42 \times 10^{-5}\ f^{1.5}\ \text{dB/yd},$$

where f is in kcs.

2.5 REFRACTION LOSS, φ [also known as the Refraction Anomaly].

If acoustic energy is propagated in a homogeneous, isovelocity ocean (so that the velocity is independent of depth), the path of the transmitted acoustic energy will be a straight line, and changes in intensity can be accurately predicted in terms of spreading and absorption loss. If the ocean is not

---

[1] W.H. Thorp, "Deep-Ocean Sound Attenuation in the Sub-and Low-Kilocycle-per-Second Region", J. Acoust. Soc. Am., 38, pp 648-654 (October 1965).

isotropic, that is, if the velocity is not constant throughout the ocean, then the energy path will be refracted from a straight line. Acoustic energy lost from the channel or from degradation of the beam by nonhomogeneous refraction gives rise to refraction losses. This anomalous behavior can be either positive or negative owing to focusing or diverging refraction effects. Energy lost from the surface channel via refraction is called "leakage" loss.

The magnitude of φ is a complicated function of the velocity geometry in the ocean in the direction of the transmitted ray. This term is one of the most difficult in the sonar range equation to evaluate because it is almost impossible to obtain an operational velocity profile that is relevant to the particular ray under study at a given instant.

The term φ , in dB, is measured (or calculated) as the logarithm of the ratio of the ~~predicted intensity (I_p)~~ measured or actual intensity (I_m) to the ~~measured or actual~~ predicted intensity ~~(I_m)~~ (I_p) at some point far from the transducer and is given by

$$\varphi = 10 \log \frac{I_m}{I_p} \qquad (2\text{-}12)$$

Transmission loss tends to start deviating from the spherical spreading loss curve at about 3500 yards. In practice, deviation from spherical spreading (-20 dB/decade of range) is observable at a range where the intensity is about 70 dB down. Any deviation from the spreading and attenuation loss curves (added together) is caused by transmission anomalies or refraction losses.

2.6 SOURCE LEVEL, S

The source level is the intensity of sound emitted by the source (whether a transducer or ship noise) at a reference distance of one yard from the effective acoustic center, expressed

as a spectrum level in dB relative to 1 dyne/cm^2 in a 1 cps band.

The value can be calculated as follows:

$$S = 71.6 + 10 \log [P] + \Delta_T \text{ dB} \qquad (2\text{-}13)$$

71.6 = Factor (dB) to convert radiated power in watts to intensity reference - the intensity with p=1 dyne/cm^2 = 1 μ bar.

[P] = Total radiated acoustic power in watts

Δ_T = Transmitting directivity index.

If the intensity of sound is measured at a known range R_o,

$$I = \frac{p^2}{\rho_o c} \qquad (2\text{-}14)$$

p = effective acoustic pressure

ρ_o = density of medium

c = velocity of sound in the medium.

The total power can be calculated from

$$P = I \cdot A = \frac{p^2}{\rho_o c} \cdot 4\pi R_o^2 \ , \qquad (2\text{-}15)$$

if the spreading is known to be spherical [attenuation is neglected].

2.7 TARGET STRENGTH, T

Target strength is similar in nature to source level except that the acoustic energy associated with the target (in the echo ranging problem) arises from scattering of incident acoustic energy and not self-radiated energy.

The target strength is defined as the ratio of the reflected intensity referred to 1 yard from the effective acoustic center, to incident acoustic intensity:

$$T = 10 \log \frac{I_1}{I_o} \qquad (2\text{-}16)$$

where

I_o = incident acoustic intensity
I_1 = reflected sound intensity referred to 1 yard from the effective reflection center.

Typical submarine target strength values are around 20 dB (100:1 I_1/I_o ratio). How is this possible? At great ranges, even though a submarine is a large object, the target appears to be a point source. The acoustic energy appears to come from a point that is known as the acoustic center of the target. The strength of the point source (a hypothetical point) is given by the intensity at a one yard reference range and the point is located several yards from the actual scattering surface and within the scatterer. It is therefore not surprising that such a large value of T is observed. This is shown by

$$I_s = \frac{I_o \sigma}{4\pi r^2} \quad , \qquad (2\text{-}17)$$

where

I_s = scattered sound intensity
I_o = incident sound intensity
σ = target scattering cross-section (including target reflection coefficient)
r = range

For the special case of r = 1 yard,

$$I_1 = \frac{I_o \sigma}{4\pi} \quad .$$

Then

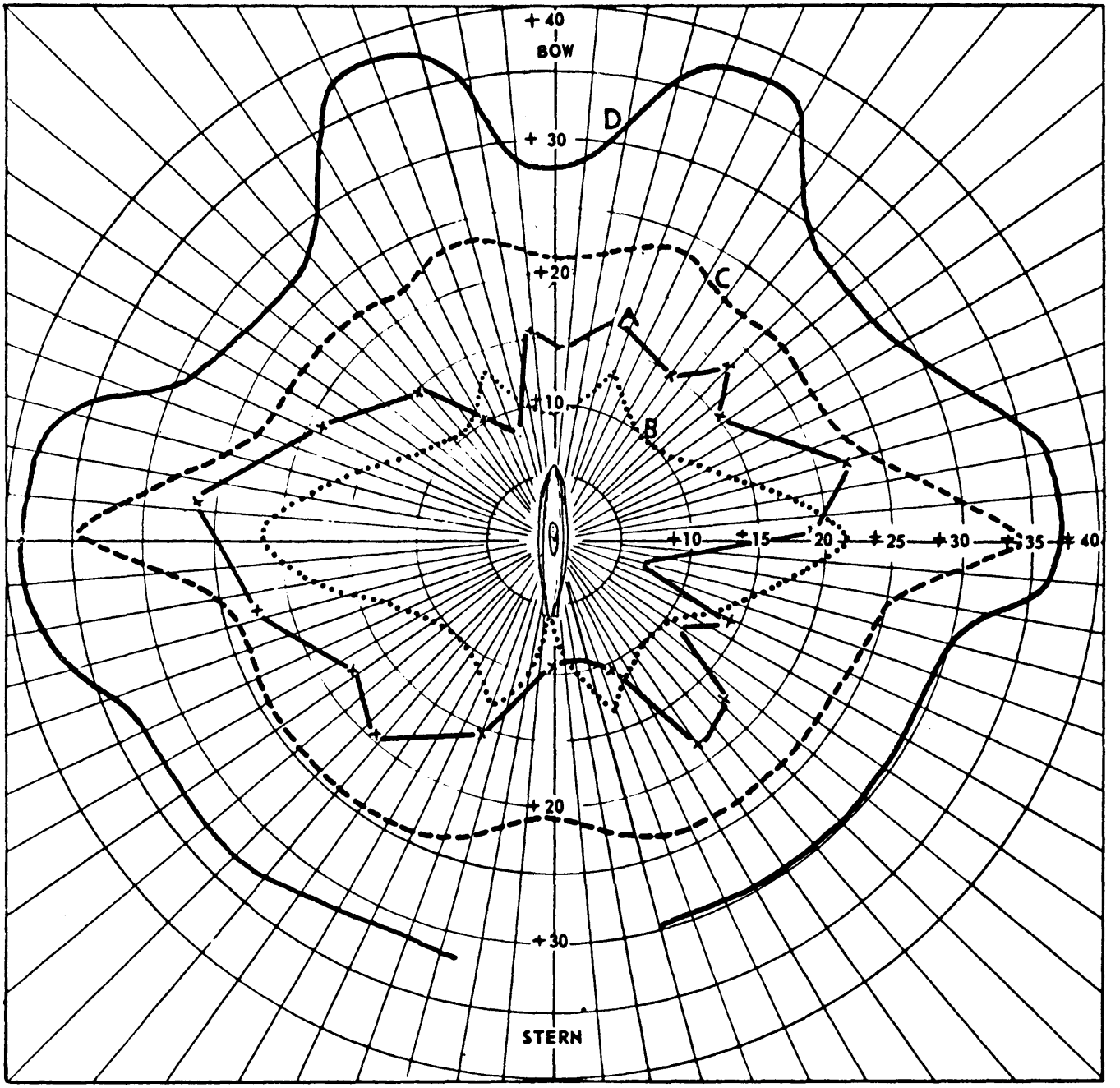
$$T = 10 \log \frac{I_1}{I_o} = 10 \log \left(\frac{\sigma}{4\pi}\right) .$$

Example: For a perfectly reflecting sphere, $\sigma = \frac{\pi d^2}{4}$ where d is the diameter.

$$T = 10 \log \frac{I_1}{I_o} = 10 \log\left(\frac{\sigma}{4\pi}\right) = 10 \log \frac{d^2}{16} = 20 \log \left(\frac{d}{4}\right) \text{dB}. \tag{2-18}$$

The scattering cross-section and the geometrical cross-section of simple, rigid, convex bodies are usually about equal, except for the effect of the reflection coefficient. The approximation is not good for complicated structures.

For a target as complex as a submarine, the usual way of determining target strength is to measure the target strength at sea, using the sonar equation, and allowing for all the medium and transmission effects. This generally leads to some average value with quite a large variation [20 dB ± 10 dB]. Fig. 2-1



A	HULL A	24kc	30ms
B	HULL B	24kc	25ms
C	HULL C	23.6kc	30ms
D	HULL A	10kc	100ms

FIG.2-1-TARGET STRENGTH VS ASPECT ANGLE

shows the relation of target strength to aspect angle. At close range, the target strength is a function of range, whereas at long range, the target strength is a function of the medium. T also varies with ping length, as shown in Fig. 2-2.

For passive sonar, the noise radiated by the target is, in a sense, its target strength. Radiated noise levels for a number of different surface ship types are shown in Fig. 2-3.

2.8 NOISE SPECTRUM LEVEL, n

The Knudsen curves (see Fig. 5-1) have long been accepted as the standard curves showing the relation of ambient noise in the ocean to sea state and acoustic frequency. They also show the absolute upper limit to the sensitivity of measuring equipment. The thermal noise level in dB relative to 1 dyne/cm^2 in a one-cycle band at a frequency f in kcps is given approximately by

$$-115 + 20 \log f \ .$$

Thus, the non-directional (isotropic) noise spectrum increases with frequency. However, when measured with a directional hydrophone, the expression becomes

$$-115 + 20 \log f - \Delta_R \ , \qquad (2\text{-}19)$$

where Δ_R is the directivity index for the receiving hydrophone.

In a receiving hydrophone, the bandwidth, w, is

$$w = \frac{f_o}{Q} \ ,$$

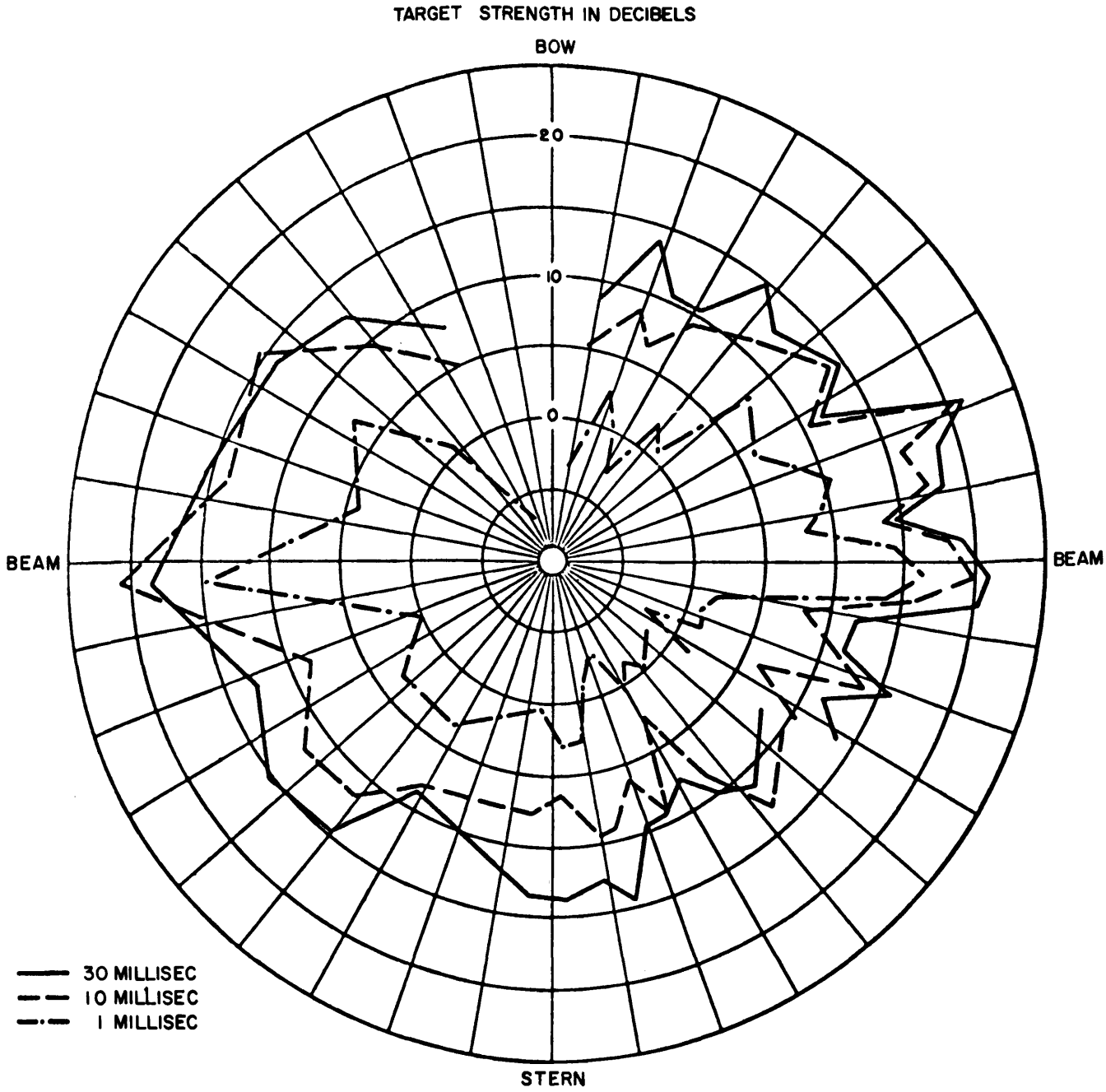


FIG. 2-2-DEPENDENCE OF TARGET STRENGTH ON SIGNAL LENGTH FOR AN S-BOAT AT 60 KC.(PART III PHYSICS OF SOUND IN THE SEA, P. 407, DIV. 6 VOL. NDRC TECH. REPTS.)

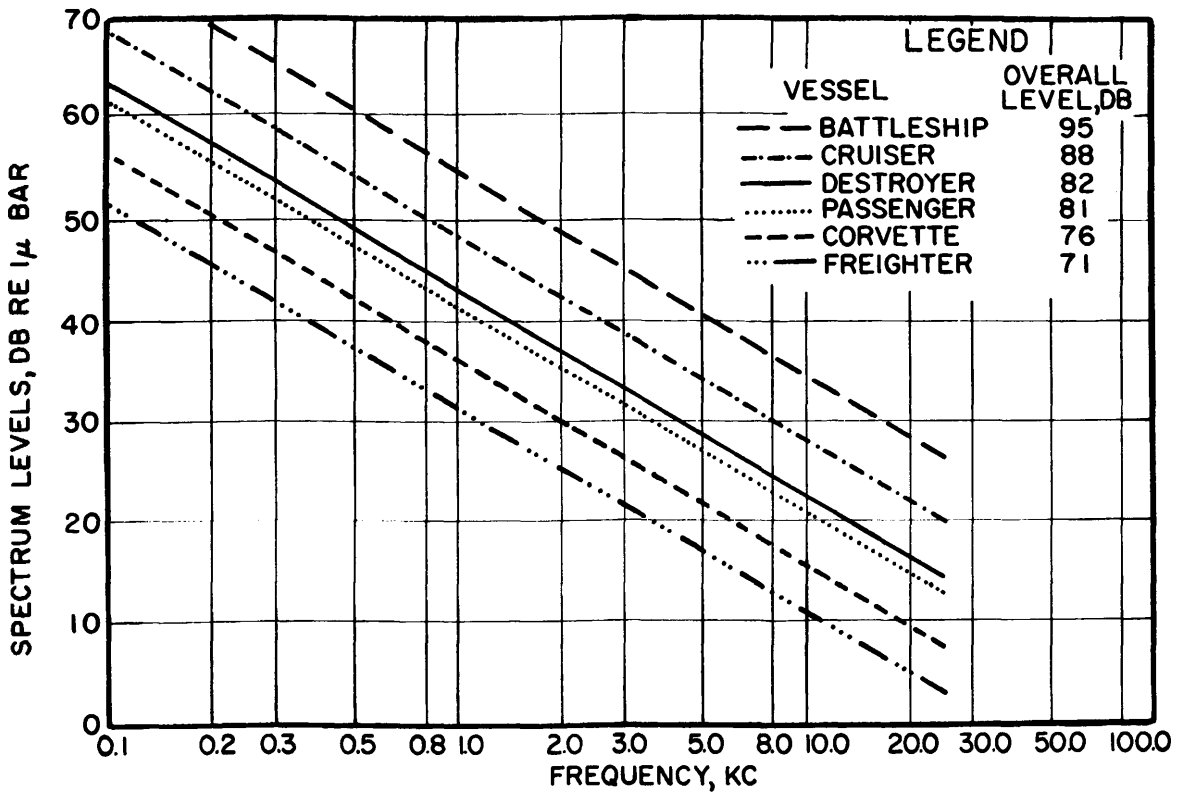


FIG. 2-3 – AVERAGE RADIATED NOISE SPECTRA OF SIX DIFFERENT CLASSES OF SHIPS

SOURCE: NDRC DIV 6, VOL 7, PRINCIPLES AND APPLICATIONS OF UNDERWATER SOUND, PAGE 236

TRACOR, INC. | DWG. 6342
AUSTIN, TEXAS 11/18/65 AG/SL

where Q is the ratio of the stored energy to the dissipated energy per cycle. If noise is assumed white, the noise output of the receiver is proportional to w; i.e.[2],

$$[N] = I(f_o)w \tag{2-20}$$

where [N] is the receiver output and $I(f_o)$ is characteristic of a particular noise ($I(f_o)$ is the intensity of the noise in a 1-cps band). Then,

$$N = 10 \log [N] = 10 \log I(f_o) + 10 \log w. \tag{2-21}$$

The function $n = 10 \log I(f_o)$ is called the spectrum level of the noise at f_o.

In the sonar equation, the noise term must include ocean noise and self noise. Spectrum levels for ship self noise are shown in Fig. 4-7. When the receiving directivity index is added, the preceding equation becomes

$$10 \log [N] = n + 10 \log w + \Delta_R. \tag{2-22}$$

A survey of all sources of ambient ocean noise in the range from 1 cps to 10^5 cps has been published by Wenz[3].

---

[2] [N] is the average noise power in watts at the transducer output. N is the corresponding dB level.

[3] G.M. Wenz, "Acoustic Ambient Noise in the Ocean; Spectra and Sources", J. Acoust. Soc. Am., 34, 1936-1956 (1962).

2.9 DIRECTIVITY INDEX, Δ

The transmitting directivity index is expressed as the ratio, in dB, of the intensity I_A at a distant point on the axis of the transducer (used as a transmitter) to that I_o of an omnidirectional projector with the same total radiated acoustic power.

$$\Delta_T = + 10 \log \frac{I_A}{I_o} \qquad I_o = \frac{P_o}{A} = \frac{P_o}{4\pi R^2} \; ; \; I_A = \frac{p^2}{\rho_o c}$$

$$= + 10 \log \frac{p^2}{\rho_o c} \cdot \frac{4\pi R^2}{P_o}$$

$$= + 10 \log \frac{4\pi p^2 R^2}{\rho_o c P_o} ,$$

where

R = range of measurement
P_o = total power radiated
p = pressure level
ρ_o = density of medium
c = sound velocity in the medium.

The receiving directivity index Δ_R for the same transducer is the negative of Δ_T, and is a negative number. Highly directional transducers have indices of the order of 30 dB. The directivity index is different, in most cases, for reverberation and noise, because noise is isotropic, and reverberation is not.

It can be shown that for a circular transducer whose diameter is greater than two wavelengths

$$\Delta_R = 20 \log \left(\frac{c}{\pi f d}\right) = 20 \log \left(\frac{\lambda}{\pi d}\right)$$

where

c = sound velocity d = transducer diameter

f = frequency λ = wavelength

Similar equations can be written for other aperture shapes.

2.10 RECOGNITION DIFFERENTIAL FOR NOISE, M_n

Recognition differential is defined as the ratio, expressed in dB, of the signal level to background noise level measured in the water near the array, required for an 0.5 detection probability. Other probabilities are sometimes chosen for appropriate reasons but are so stated. The smaller this number, the better the over-all system performance. This quantity traditionally has involved the performance of a human observer. It is therefore in part a psychophysical factor depending upon the operator [alertness, fatique, mental outlook]. When automated detection systems are considered, a "decision rule" for detection is involved.

The usual discussions of recognition differential do not consider "clutter" rate or "false alarm" rate. This failure has been one of the primary shortcomings of the concept. An example will be given in Chapter 6 of two processors with identical output signal-to-noise ratios which cannot provide the same detection performance.

RECOGNITION DIFFERENTIAL FOR REVERBERATION, M_R

This quantity is defined the same as M_n except that the total noise term [n + 10 log w + Δ_R] in Eq. (2-1) is replaced by the reverberation level Q, as in Eq. (2-2).

2.11 REVERBERATION LEVEL, Q

The reverberation level replaces the total noise term in the echo ranging equation [if the maximum range is reverberation limited]. Reverberation level is defined as the ratio, in dB, of the scattered sound intensity (I_R) at the array to a reference intensity I_o, i.e.,

$$Q = 10 \log \frac{I_R}{I_o} ,$$

where I_o is usually the intensity associated with an acoustic pressure of 1 dyne/cm^2.

2.11.1 Volume Reverberation. A random distribution consisting of x small scatterers per unit volume and an active volume V will provide a reverberation intensity,

$$I_R = \frac{I_1 \, xV \, \sigma}{4\pi \, r^4} ,$$

where

I_R = reverberation intensity

I_1 = source intensity

σ = scattering strength per scatterer

r = range (with transmitter and receiver at same distance from V).

The volume V of scatterers producing return simultaneously is

$$V = \frac{\alpha}{4} \; r_o \, r^2$$

where

V = active volume

r = range to center of volume

r_o = ping length = $c\tau/2$

α = horizontal angular width of transducer beam (radians).

β = vertical angular width of transducer beam (radians).

$$\therefore \quad I_R = \frac{I_1 \ \pi\alpha\beta r_o x\sigma}{2r^2}$$

The reverberation intensity is proportional to

a. source intensity - I_1

b. ping length - r_o

c. beamwidth - α, β

The volume reverberation intensity is inversely proportional to the range squared.

2.11.2 Surface Reverberation (Including Bottom): The sound intensity scattered from an area rather than a volume has a different behavior.

The reverberation is given in general by

$$I_R = \frac{I_1 \ x \ A\sigma}{4\pi r^4} \quad .$$

In this expression, x is the number of scatterers per unit area, and $a \cong 2\alpha r \ r_o$ if the projector is close to the surface.

At great ranges,

$$I_R \cong \frac{I_1 \times \sigma r_o \alpha}{2\pi r^3} .$$

Surface (or bottom) reverberation varies inversely as the third power of the range.

2.11.3 Characteristics of Reverberation. When operating in low sea states with a quiet ship, reverberation is usually the predominant term in the background waveform. A number of statements can be made concerning the reverberation.

1. In surface channel operations, the early portions of the echo cycle are dominated by reverberation (surface) which decreases as $1/r^3$. In the latter portions of the echo cycle (range 20,000 yds), reverberation (volume) is decreasing as $1/r^2$. This statement must be modified if the water is shallow and bottom reverberation effects become involved.
2. Surface reverberation is very likely to be the dominant background at the time a target is likely to appear when operating in the bottom bounce or convergence zone modes.
3. Since reverberation originates from a large number of individual, randomly-located scatterers, the reverberation amplitude can be described with the Rayleigh distribution.

2.12 CRITIQUE OF RECOGNITION DIFFERENTIAL

The recognition differential is usually stated as that signal-to-noise ratio at a specified place in the system which is necessary for an alert observer to detect 50% of the targets. The recognition differential, however, has a serious failing in

that the percentage of correct calls, "hits", is usually specified without also specifying something fixed about either false alarms or correct rejections. This state of affairs came about because implicit assumptions were made about an observer's performance which were not true. Let us make some of the assumptions clear, show how they affect the recognition differential, and state alternative assumptions with new psychophysical procedures which circumvent the failing mentioned.

The typical situation has an observation interval (whether vision or hearing is used is immaterial) in which there is noise or noise plus signal. The observer's task is to examine the stimulus material and say "yes" or "no". He says "yes" if he detects a signal, and "no" if he does not. To measure the recognition differential, the signal-to-noise ratio is varied and we generate a curve such as that in Fig. 2-4 where the abscissa is signal-to-noise ratio and the ordinate is percentage of correct detections, P(Y|SN), i.e., the probability of saying "yes" when signal-plus-noise was in the observation interval. This value has been determined by a variety of psychophysical methods and is commonly called the masked threshold.

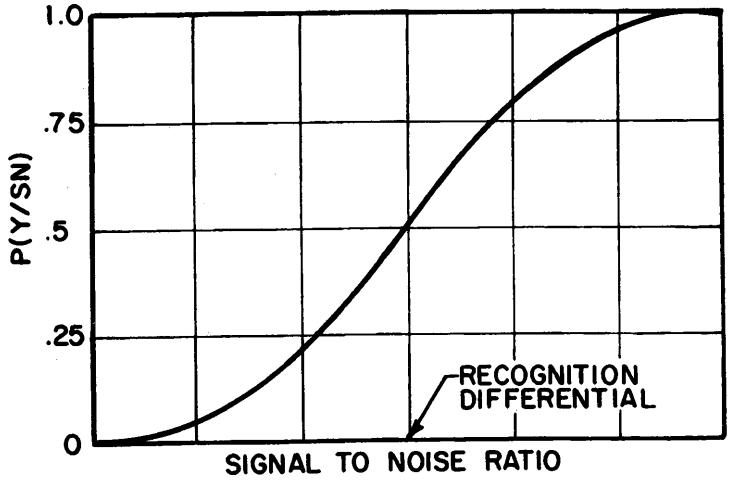


FIG. 2-4 - DETECTION PERFORMANCE

What is not controlled in finding this value, however, constitutes a very serious failing as we have remarked earlier, because if we ask the observer to adopt a very strict criterion ("Be very sure that a signal is present.") the whole curve is shifted to the right and the recognition differential (the signal-to-noise ratio for 50% detections) is higher. On the other hand, if we require the observer to be very lax in his criterion, ("Be very sure not to miss any signals.") the whole curve is shifted to the left. In the first instance, he misses more often; in the second instance, he false-alarms more often. The question then is whether we can get a criterion-free measure so that we can make a meaningful statement about an observer's performance in signal detection. The answer is "yes". Let us now show why "yes" is the proper answer.

2.12.1 The Sonar Detection Problem. The detection problem for a shipboard sonar operator is one of finding a signal within a masking noise background, In finding this signal, he searches for one of a set of peculiar combinations of characteristics that, for him, denote a signal. The quantity, "signal", is a multi-dimensional quantity with dimensions such as loudness, duration, center frequency, bandwidth, frequency spectrum, and modulation for audio detection, and brightness, extent in range, extent in azimuth, and structure for video.

The masking background noise within which he searches for the signal has components of all the above characteristics. The operator's task consists of establishing a weighted average of all the dimensions when no signal is present, and establishing some minimum deviation from this weighted average which he will interpret as the presence of a signal.

If we knew what weighting factors should be assigned to each component, the process of detection could be described in concrete, specific terms. At present, however, the interactions among the variables are not known. Some progress in

analysis and description of the detection problem can be made, despite our lack of specific knowledge. Experiment has shown, for example, that if a large number of samples of noise are observed, the effect of the individual components, combined in a manner we cannot describe, can be rated according to a single scale. This scale we can call receiver response. For a particular set of total environmental conditions, a large number of samples of noise, rated in terms of receiver response, will be found to be distributed about some particular value of response.

Similarly, for any particular set of total environmental conditions, when a large number of signals are observed within the noise and rated in terms of response, they also will be found to be distributed about some particular value of response.

Suppose many samples of noise and signal plus noise are taken under the same conditions and plotted against an arbitrary response scale. For many operational situations, the resulting distributions will be Gaussian and will have the same variance. This phenomenon is familiar to workers in underwater acoustics, who, in making measurements involving sound pressure levels, generally expect that for a large number of samples, the standard deviation will be about 6 dB. Although the two distributions under discussion have the same variance, their mean values will be different. Since the response scale is arbitrary, we can choose any convenient measure for it. Because the scale is subjective, and position along it results from an indeterminate combination of many factors, there is no concrete physical unit or relationship available that we can use as a scale unit. We therefore turn to the distributions themselves for a unit and choose their standard deviation as the unit measure along the response axis. This not only provides a self-contained unit measure, but also serves to generalize our plot of noise and signal plus noise distributions.

Generalized diagrams formed in this manner provide a tool for study of detection performance, a tool which can be useful despite our lack of precise knowledge of the processes involved in interpreting and recognizing noise and signal plus noise. In addition, they permit the comparison of performance under highly diverse conditions for which comparison would otherwise be difficult or impossible.

2.12.2 The Theory of Signal Detectability. In the analysis of a sonar system, the response of the operator to the information presented on the sonar display determines to a large extent the overall effectiveness of the system. Ideally, investigations of operator performance would be carried out at sea, using trained sonar operators on an operational sonar. In practice, this sort of experiment is not as productive as it appears. It is very time-consuming and costly, and there is so little control over the numerous variables influencing the test that the results are difficult to interpret, and restricted to a few special cases.

Controlled tests in a laboratory environment, while seemingly remote from any application to a realistic shipboard detection situation, do provide a means of measuring human detection performance and relating these measures to the performance of a shipboard sonar operator. The unifying theory, the Theory of Signal Detectability (TSD), has been developed since 1954[4]. TSD is now a well worked out body of theory which relates statistical decision theory to the general detection problem. Although the theory was originally conceived to describe mathematically ideal or optimal detection processes, it

---

[4]W.W. Peterson, T.G. Birdsall, and W.C. Fox, "The Theory of Signal Detectability", IRE, Trans. Inform. Theory 4, 171-212 (1954).

has become apparent that it represents a good approximation to a descriptive theory of human detection and recognition behavior[5]. It has also been developed, especially in the context of hearing, in such a way as to relate it more closely to stimulus parameters[6]. For the time being, we will confine our discussion to the ideal case.

The fundamental detection problem of TSD introduces and defines two concepts, the index of detectability, d', and the criterion, c, which provide useful measures of performance. To arrive at these two quantities, we construct a plot of the probability density functions for noise alone, N, and for signal plus noise, SN, in the same manner as described in Section 2.12.1. We assume that the two plots are Gaussian, and have the same variance, and that the mean value of SN is larger than the mean value of N. The resulting plot, shown in Fig. 2-5, is two overlapping distributions.

The distance between the means may be regarded as a function of the effectiveness of the signal or alternatively as a measure of the observer's sensitivity. This distance, expressed in the units of standard deviation previously selected as a scale, is the index of detectability, d'. As a measure of the difference in sensory response to noise and signal plus noise stimuli, it is apparent that d' is analogous to a signal-to-noise ratio. At the same time, it is also apparent that, despite this close analogy, it is impossible with our present knowledge to assign definite values of signal-to-noise ratio to d', except in special cases, where the situation can be limited to a single dimension.

---

[5] J. A. Swets, Ed., "Signal Detection and Recognition by Human Observers", New York, John Wiley and Sons, 1964, p.v.

[6] Jeffress, L.A., "Stimulus-Oriented Approach to Detection", J. Acoust. Soc. Am. 36, 766-774 (1964).

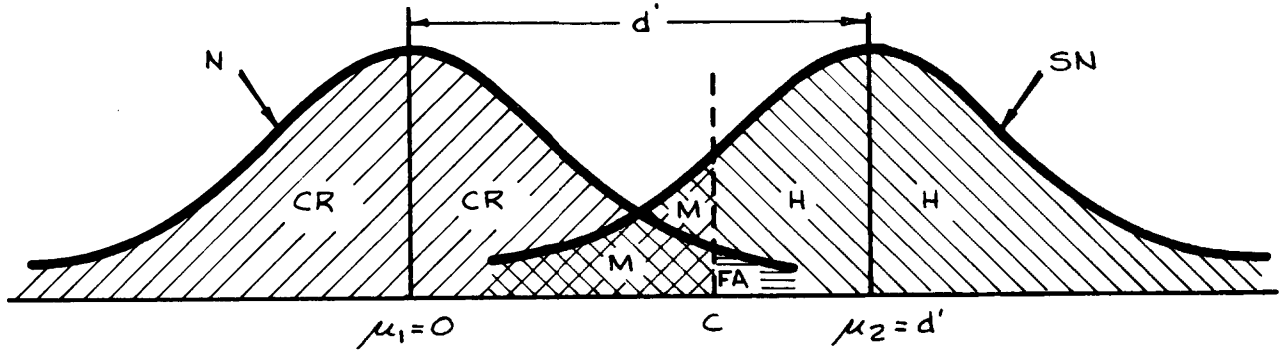


FIG. 2-5 OVERLAPPING NORMAL DISTRIBUTIONS FOR OBTAINING d'

TRACOR, INC. | DWG. 734-95
AUSTIN, TEXAS 7-5-65 BHD-MM

The mean of the N distribution in Fig. 2-5 has been placed at zero so that d' is read directly as the position along the abscissa of the mean of the SN distribution.

In an experiment, an observer is provided a series of fixed time observation periods. At the end of each period, he reports his observation as noise only, or as signal plus noise. Thus, his task during each observation period is to decide whether the observed sample came from the N or the SN distribution. We assume that to accomplish this task, the observer establishes a criterion, represented by some point, c, along the abscissa of Fig. 2-5. He responds "no" when a particular observation is less than c, and "yes" when it is greater.

Just as d' was analogous to signal-to-noise ratio, c can be seen to be analogous to a threshold. And, for the same reasons as in the case of d', no definite value can be assigned to c.

Of the two parameters described above, d' is fixed by the experimental conditions and defines the limitations imposed on the observer by the characteristics of N and SN, and by his own sensitivity to these characteristics. The observer, however, to some degree controls c. His control in an operational situation may be unconscious or intuitive. In the laboratory, the observer is made to control c by the instructions he is given at the start of the experiment ("Don't miss any signals", or "Be very sure you have a signal when you say yes.", for example).

The familiar Receiver Operating Characteristic (ROC) curve can be generated directly from the N and SN distributions of Fig. 2-5. The area under the N distribution to the right of the criterion represents the probability of false alarm, the probability of saying "yes" to noise alone, p(Y/N), for that criterion. Similarly, the area under the SN distribution to the

right of the criterion represents the probability of detection, p(Y/SN), for the same criterion. Each possible criterion generates a pair of values which may be mapped as a point in the p(Y/SN) - p(Y/N) plane (See Fig. 2-6). The smooth curve resulting from mapping every such point is the ROC curve. When the variances of the SN and N distributions are equal, as has been assumed here, the ROC curve will be symmetrical about the negative diagonal of the plane, criterion is a position along the ROC curve, and d' is the distance along the negative diagonal between the ROC curve and the chance line. Thus, a set of values of d' defines a family of ROC curves. Further, under the assumptions that have been used so far, a ROC curve is completely described by the value of d'.

In practice, there are several ways of obtaining d' for a given observer in a psychophysical experiment. The simplest is merely to present the observer with a series of trials in which a signal may or may not be present on each trial. The results from such an experiment are then collected into a four-fold table of response proportions representing estimates of probability of a hit, P(Y|N+S); probability of false alarm, P(Y|N); probability of a miss, P(NO|N+S); and the probability of a correct rejection, P(NO|N). These values may then be used to enter a table of areas under a unit area normal distribution to obtain d'. Such an experiment, however, yields only the one point on an operating characteristic on the negative diagonal. Since the whole ROC provides important additional information, it is desirable to use a method which allows several points to be plotted and an ROC fitted to these. Such a method, called the rating scale method, is made possible by giving the observer more than two response alternatives. Actually, this amounts to letting the observer adopt several criteria simultaneously. This is done by asking the observer to rate each period during which a signal might have been presented on, say, a four-point rating

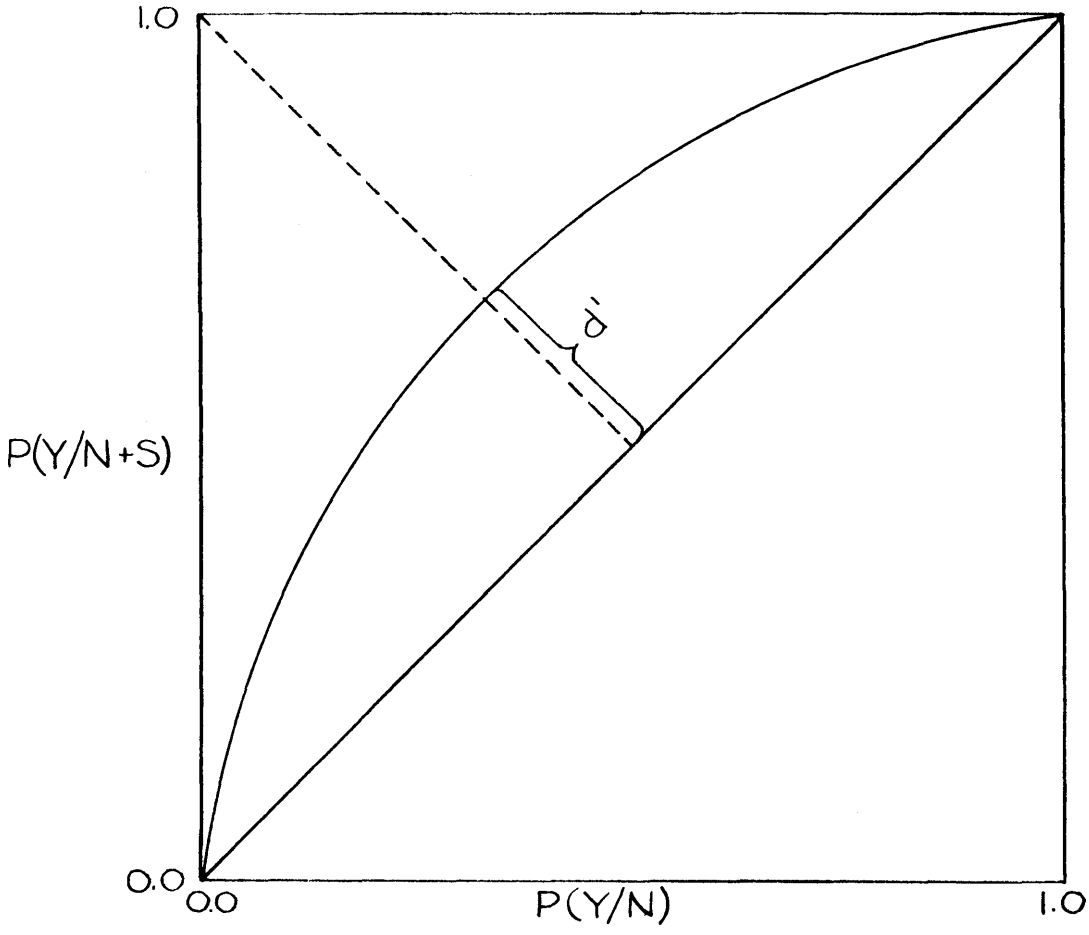


FIG. 2-6 - RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE

TRACOR, INC. | DWG. 771-25
AUSTIN, TEXAS | 5-26-65 BMB-ES

scale: "1" meaning that he has extreme confidence that a signal was present, "4" meaning that he is very confident that no signal was presented, and "2" and "3" indicating intermediate confidence ratings. Such a four-point rating scale gives rise to three points on an ROC. If these are plotted on normal-normal paper, a best-fit line may be drawn and d' obtained from the graph.

2.12.3 Modifications. We have discussed the ideal case rather fully in order to establish a basis for a discussion which will relate actual psychometric data more nearly perfectly to underlying physical processes. Jeffress[6] has shown how, when a psychophysical experiment is performed which will generate a ROC, the shape typically deviates from the curve shown in Fig. 2-6. In Fig. 2-6, the curve is drawn symmetrically about the negative diagonal. Experimentally generated curves, however, bulge somewhat away from the chance line, or positive diagonal, in the half below the negative diagonal, and approach a little closer to the positive diagonal above it. Without stating Jeffress's reasoning fully, we will merely state that this new shape can be approximated very closely by assuming that N has a Rayleigh distribution and, when signal is added, the resulting signal plus noise has a Marcum distribution (See Sect. 6.4.3, et seq.) Again, without going into details, the outcome of using these distributions alters the exact definition of the detectability index to a new one designated d_s. The single number, d_s, is useful in describing the detectability of signals in an over-all way in that it will also imply the whole shape of the ROC. It will not specify the particular behavior of an operator in a situation because it is a function of many unspecified variables such as his unconscious setting of criterion, c, in accordance with implicit values and costs. That is to say, for the very same physical signal-to-noise ratio different operators will have results that place them in different places on a ROC, and, indeed, different operators may behave as if they were working with a different signal-to-

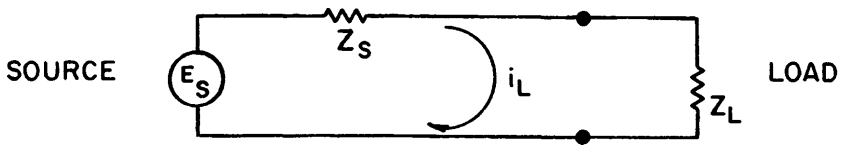
noise ratio, because they produce differing d_s. One can only speculate that this means that different operators have different amounts of "internal noise", or different effective bandwidths.

3. TRANSDUCERS

In this part we will start with some fundamental notions of analysis and performance for various types of transducers. For the time being, "transducer" will mean an element which converts an electrical signal into an acoustical signal or vice versa. Later, transducer will mean, in a larger sense, an array or large collection of such elements. Details concerning the behavior of arrays will be given in Chapter 4.

3.1 ANALOG REPRESENTATION OF DYNAMICAL SYSTEMS

Let us begin by considering a simple single frequency energy transmission system. A power source furnishes power to a



where e_s - open circuit source voltage

z_s - source impedance

i_L - load current

z_L - load impedance

The load current is then given by

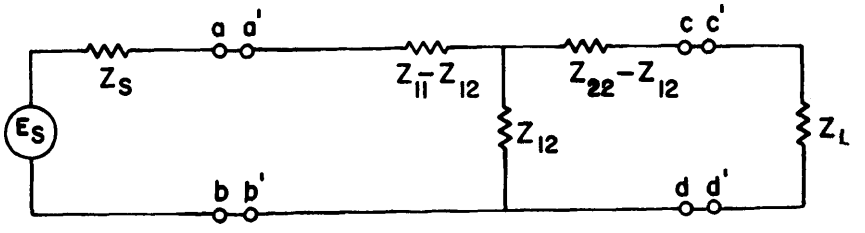
$$i_L = \frac{e_s}{z_s+z_L} \; ; \tag{3-1}$$

and the power P_L transferred to the load is $i_L^2 R$ where R is the resistive component of the load.

$$P_L = e_s^2 \frac{R_L}{|z_s+z_L|^2} \tag{3-2}$$

The impedances are, in general, complex. For a given e_s, the

maximum power is transferred to the load if $R_L=R_s$ and $X_L=-X_s$, i.e. if the source and load have conjugate impedances. Now suppose that our problem is to transmit an acoustic signal in water. The impedance of the water represents a load to which the energy source (power amplifier) must supply energy. In this case we need a transducer to complete the operation. The characteristics of the transducer must be such that energy is transferred from the source terminals a and b (below) to the water, represented in analog form by z_L with terminals c' and d'. The transducer is the T network between terminals a' and b' and terminals c and d. For maximum power transfer to the water the source impedance z_s must match the impedance looking into terminal a' and b' with z_L connected. A second requirement is that the impedance looking toward the source from terminals c and d must match z_L. When these conditions are satisfied, the most efficient transfer of energy from the source to acoustic from in the water takes place.



To this point we have represented electrical, mechanical and acoustical impedances as equivalent electrical circuits for a system. This representation comes under the subject of dynamic analogies. This subject is a well developed technology. The electrical circuit analog to a mechanical or acoustical system provides a convenient representation for the differential equations which describe the real system. The formal solution of the equivalent analog is easily carried out by well known elementary formalisms, whereas simplifying the differential equations and obtaining results in terms of the dynamical system has not been so formalized.

All electrical analogies used to represent dynamical systems depend upon the similarity of the functional form of the ratio of potential difference to current and upon ratios like force to velocity or pressure to particle velocity or volume flow. In general, two types of analogies are employed: those referred to as impedance or "force-voltage" analogies in which potential difference represents force or pressure and those referred to as the electromechanical or "force-current" analogies. Either can be applied in most situations. The "force-voltage" analog will be used in this discussion.

A simple example will illustrate the use of an analogy in determining the behavior of a mechanical system. The mechanical system is shown in Fig. 3-1(a). It consists of two masses M_1 and M_2 connected by a spring immersed in a fluid. Forces F_1 and F_2, exerted at points a and b, drive the system. Viscous fluid action on the masses is assumed to be present. Gravitational effects are assumed to be entirely absent in this example. The corresponding mechanical schematic is shown in Fig. 3-1(b). The only difference between diagrams (a) and (b) is the addition of dashpots to represent viscous forces. If the positions of the points c and d are x_1 and x_2 relative to the origin of coordinates, the tension T in the spring is

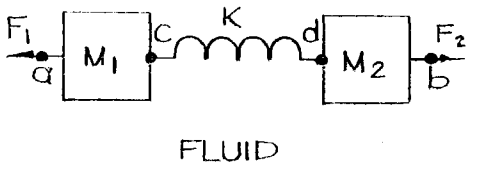
$$T = K(x_1 - x_2) \quad . \tag{3-3}$$

The forces F_c and F_d applied by the dashpots to the points c and d are

$$F_c = H\dot{x}_1 \quad \text{and} \quad F_d = H\dot{x}_2 \quad , \tag{3-4}$$

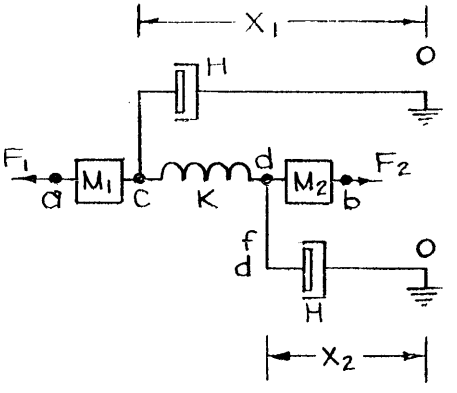
where the "dots" denote differentiation with respect to time; i.e., the velocities of the points c and d. The sum of the forces acting on M_1 is the rate of change of momentum of M_1, so

$$F_1 - F_c - T = M_1\ddot{x}_1 \tag{3-5a}$$

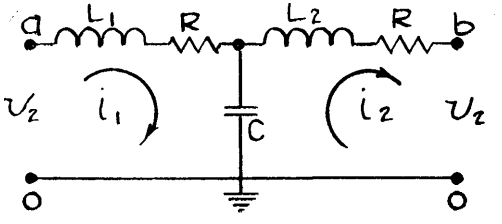


FLUID

ACTUAL SYSTEM
(a)



MECHANICAL SCHEMATIC
(b)



ELECTRICAL SCHEMATIC
(c)

Fig.3-1-ELECTRICAL ANALOG TO A DYNAMICAL SYSTEM

TRACOR, INC. | DWG. 771-30
AUSTIN, TEXAS | 5-27-65 BMB-ES

and the corresponding expression for M_2 is

$$T - F_d - F_2 = M_2\ddot{x}_2 \tag{3-5b}$$

Equations (3-3), (3-4), and (3-5) describe the complete behavior of the spring, the dashpots, and the masses. The parameters K, H, M_1, and M_2 determine the magnitudes of their behavior characteristics, each being a ratio of a force to a velocity, $\dot{x}$, a force to the derivative of a velocity, $\ddot{x}$, or a force to the integral of a velocity, x. These quantities may be considered to be completely analogous to the quantities resistance, R, inductance, L, and capacitance, C, used in describing electrical circuits, because these quantities are the ratios of a potential difference and a current, i, derivative of a current, di/dt, or an integral of a current. Elimination of F_c, F_d, and T from Eqs. (3-5a) and (3-5b) leaves

$$\begin{aligned} F_1 - (M_1\ddot{x}_1 + H\dot{x}_1 - Kx_1) &= -Kx_2 \ , \\ F_2 + (M_2\ddot{x}_2 + H\dot{x}_2 + Kx_2) &= -Kx_1 \end{aligned} \tag{3-6}$$

The electrical circuit in Fig. 3-1(c) is an analog for the mechanical system. This can be shown by writing the circuit equations for Fig. 3-1(c),

$$\begin{aligned} v_1 - L_1\frac{di_1}{dt} - Ri_1 - \int_o^t \frac{(i_1-i_2)}{c}\,dt &= 0, \\ -v_2 - L_2\frac{di_2}{dt} - Ri_2 - \int_o^t \frac{(i_2-i_1)}{c} &= 0. \end{aligned} \tag{3-7}$$

By substituting $\dot{q}$ (charge flow rate) for i, the resulting

equations are

$$v_1 - (L_1\ddot{q}_1 + R\dot{q}_1 + \frac{1}{c}\, q_1) = -\frac{1}{c}\, q_2$$

$$v_2 + (L_2\ddot{q}_2 + R\dot{q}_2 + \frac{1}{c}\, q_2) = -\frac{1}{c}\, q_1 \quad . \tag{3-8}$$

The forms of Eqs. (3-8) and (3-6) are identical if v_1 and v_2 represent F_1 and F_2 and $\dot{q}_1$ and $\dot{q}_2$ represent x_1 and x_2, i.e. if voltages are used to represent forces. With this understanding,

L_1 is the analog of M_1,

L_2 is the analog of M_2,

R is the analog of H,

C is the analog of 1/K.

In order to determine the behavior of the mechanical system the solution to the electrical circuit can be found using impedance techniques. For steady state behavior, with periodic input v_1, the root-mean-square values of voltage and current are given by

$$V_1 = [R + j(x_1 - x_c)]\, I_1 + j\, x_c\, I_2$$

$$V_2 = -[R + j(x_2 - x_c)]\, I_2 + j\, x_c\, I_1 \quad . \tag{3-9}$$

Solution of these simultaneous equations gives an expression for V_1/I_1, the impedance looking into the network at the left, in terms of V_2/I_2, the load impedance into which the network operates. Choosing the impedance to be connected at b then determines the impedance which will be measured at a. If V_1 is given, then I_1 and the power input to the network can be found, and I_2 and the power output to the load can be found. Hence, the role of this mechanical system in Fig. 3-1(a) in transferring energy from the

apparatus (not shown) furnishing F_1 to the apparatus (also not shown on which F_2 is exerted can be completely described.

It takes a little practice to "guess" the equivalent analog for a given mechanical system. As a last resort the mechanical equations of motion can be written out in detail and transformed directly into electrical terms. The resulting equations can either be solved directly using the impedance approach or they can be used to construct the analog circuit. In general, both mechanical and acoustical behavior will appear in the analog circuit. Use of analog circuits is, of course, not essential to a solution but helps to visualize the system. The mechanical equations can be solved directly.

It will be instructive to write Eq. (3-9) in terms of a generalized nomenclature. In (3-9), $-jx_c$ is the impedance common bo both the input current I_1 and the output current I_2. If the nomenclature Z_{12} is used to represent the common impedance to the input and output currents, then

$$Z_{12} = -j\, x_c \quad . \tag{3-10}$$

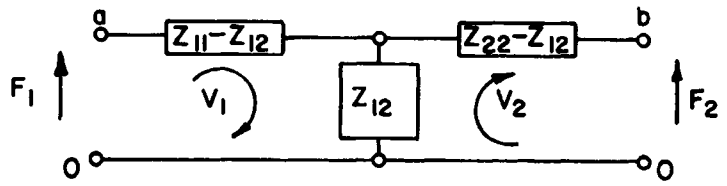
If Z_{11} and Z_{22} are considered to be the impedance seen looking into the input with no load and the impedance looking back from the output with no source,

$$\begin{aligned} Z_{11} &= R + jx_1 - jx_c \\ Z_{22} &= R + jx_2 - jx_c \quad , \end{aligned} \tag{3-11}$$

whence it follows that Eq. (3-9) can be written

$$\begin{aligned} V_1 &= Z_{11}\, I_1 + Z_{12}\, I_2 \\ V_2 &= -Z_{22}\, I_2 - Z_{12}\, I_1 . \end{aligned} \tag{3-12}$$

The equivalent circuit for the mechanical system is



This circuit is identical to the coupling network shown on page 45.

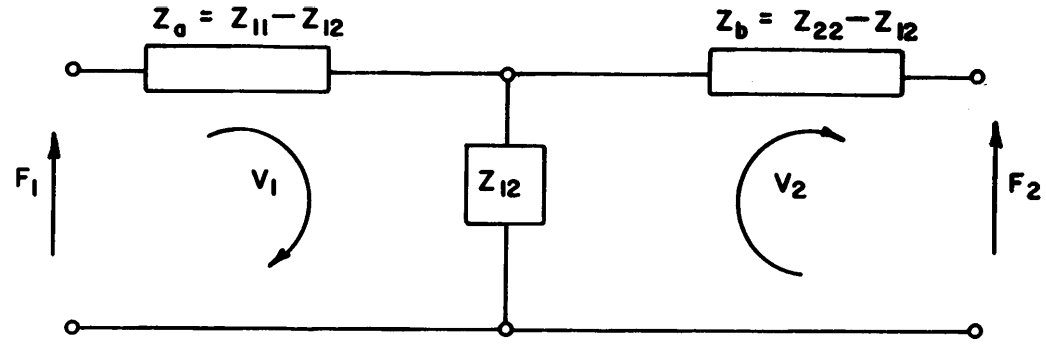
A system such as the one shown in Fig. 3-1 is frequently called a lumped parameter system because the parameters are separated into purely mass-like, spring-like and damping components. When a mechanical system cannot be described with separate masses, springs, and dashpots as was done in Fig. 3-1, that is when the springs have mass or when the masses are springs, the analysis is not so simple but results which have the same form as those presented in Eq. (3-12) can be obtained. A simple example of a distributed parameter mechanical system is the straight rod on which periodic forces F_1 and F_2 are acting at the ends. If u(x) represents the displacement at a point x in the rod, then the actual motion must satisfy the differential equation

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \quad , \qquad (3\text{-}13)$$

where x is the coordinate describing position in the rod, t is time and c is the speed of sound in the rod.

The solution of this equation in terms of the force-to-velocity ratio at each end can be found by standard mathematical techniques; the results are identical in form to Eq. (3-12) and the analog circuit is the T network above. The only difference between the equations describing the rod and the equations describing the spring coupled masses is in the expressions for Z_{11}, Z_{12}, and Z_{22}. It should be pointed out that the behavior of the rod could be approximated by the system of Fig. 3-1(a). The results from Eq. (3-13) simply give a more precise description.

Later in this discussion a typical transducer element will be analyzed using the results described above. It will be shown that parts of the element can be treated as rods whose motion is described by Eq. (3-13). In this development a single mechanical component will have the schematic representation shown here:



and the equations used to describe it will be

$$F_1 = (Z_a + Z_{12})v_1 + Z_{12}\, v_2 \qquad (3\text{-}14)$$

$$F_2 = -(Z_b + Z_{12})v_2 - Z_{12}v_1 \qquad (3\text{-}15)$$

A complex mechanical system consisting of several simple components will be made up of several of these networks connected together.

3.2 TYPES OF TRANSDUCER ELEMENTS

3.2.1 Piezoelectric

The Piezoelectric effect is the electric polarization of certain crystals which are subjected to strains. The polarization is a linear function of the strain and consequently changes sign with the strain. The converse effect also holds, i.e., the application of an electric field produces a strain in the crystal, viz, $\xi \approx rs + kD$ where ξ is the deformation, D is the displacement field, s is the mechanical stress, and r and k are constants. Crystals exhibiting the piezoelectric effect include quartz and Rochelle-salt. This type

element has found wide application in sonar since World War I.

3.2.2 Electrostrictive

The electrostrictive effect, found in most materials, differs from the piezoelectric effect in that if an electric field is applied to a specimen, the specimen undergoes a deformation proportional to the square of the displacement field. This relation can be represented as

$$\xi \approx rs + kD^2 \tag{3-16}$$

Fortunately, some ceramics such as barium titanate have the property that a large polarization field can be frozen into the material. If an alternating electric field is then applied to the ceramic the deformation will be

$$\xi \approx rs + k(D+\delta)^2 \tag{3-17}$$

where δ is the incremental displacement field. If $\delta \ll D$ the incremental deformation $\xi' = \xi - kD^2$ becomes

$$\xi' \approx rs + 2kD\delta \tag{3-18}$$

and the electrostrictive ceramic follows a linear relation as in the piezoelectric case.

One great advantage of the ceramic type is that it can be molded into any desired shape. The ceramic transducer has recently been used extensively in sonar applications.

3.2.3 Magnetostrictive

The magnetostrictive effect is noted in several materials and is exhibited by a change in length upon application of a magnetic field. The direction of the change in length is

independent of the direction of the magnetic field, so that if an oscillating field is applied, the frequency of motion of the rod will be twice that of the oscillatory field. Nickel possesses the magnetostrictive effect to a greater degree than any other material.

The use of magnetostriction to produce sound waves is accomplished as follows: A rod of the magnetostrictive material is surrounded by a coil of wire. Electrical current flowing through the wire produces a magnetic field which shortens the rod. The frequency doubling mentioned above can be prevented by either giving the rod a suitable permanent magnetization, or by sending a "polarizing current" (a continuous dc) through the coils.

3.3 ANALYSIS OF A TRANSDUCER ELEMENT

When the transducer was pictured as an electrical quadrupole connecting the source to the load, a gross simplification was introduced. From a system point of view, of course, the transducer can be considered in such a light. However, it would be hard to design an effective transducer from such a simple analysis. The transducer is in fact a very complicated device of electromechanical components, the behavior of each of which must be understood in order to approach the problem of transducer design on a scientific basis. If we consider a typical "length expander" transducer we see that it is composed of many parts including the segmented ceramic, head, tail, stress rod, end caps, and case. We can, of course, attack the problem in a rather primitive fashion by considering each piece in a lumped parameter fashion. We can then draw a mechanical circuit for the transducer and subsequently an equivalent circuit. The resulting circuit is "shape-wise" a reasonable approximation to the transducer; however, the impedances in the equivalent circuit are primitive estimates of the actual system.

It should be emphasized that whether the transducer is analyzed from the lumped parameter point of view or from the more refined continuous-medium point of view, any equivalent circuit which is drawn is merely an aid in visualizing the transducer. It is not necessary to the analysis; strictly, only the various equations of motion and state are essential to the analysis. An example of a lumped parameter analysis of a typical transducer element follows.

3.4 A TYPICAL EQUIVALENT CIRCUIT[1]

Although the exact equivalent circuit of the transducer must, of course, depend on the form of the specific transducer, it will be found that many transducers are special cases of that shown in Fig. 3-2. Naturally, Fig. 3-2 is a very simplified schematic of the actual transducer, since the head and tail may have arbitrary shapes (they may be hollow for example), but it is sufficient to produce a very general equivalent circuit.

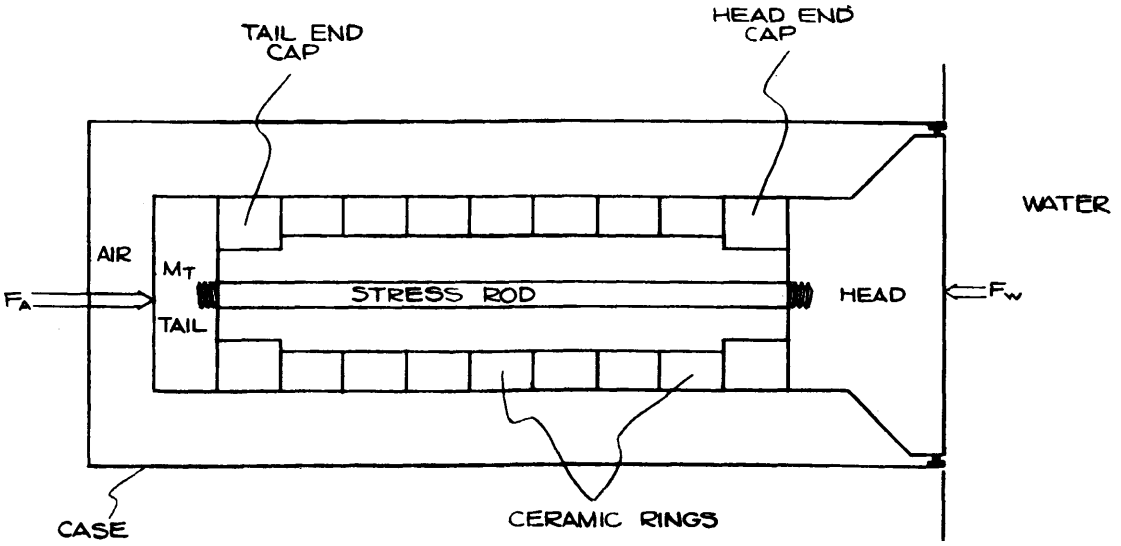


FIG. 3-2 TRANSDUCER CROSS SECTION

[1]This section is based on the work done at the U. S. Navy Electronic Laboratory by Messrs. J. Hickman, G. E. Martin, D. L. Carson.

The transducer components are a ceramic "heart," consisting of a number, n, of toroidal ceramic pieces, with toroidal end caps driving the head on one end, and driving the tail on the other end. A stress rod linking head and tail is threaded through the ceramic toroids and is attached to the head and tail. A case, supported at the head, encloses the whole apparatus. When the ceramic elements are electrically driven there is a force F_w, due to radiation, into water at the head and a force F_a into air at the tail. In general,

$$F_w >> F_a \tag{3-19}$$

As a first step, one must construct the mechanical circuit of the transducer of Fig. 3-2, and from this find its electrical equivalent circuit. To do this, its elements (head, tail, rod, etc.) may be approximated as rigid masses connected by springs (and dashpots, if it is necessary to take losses into account) as described in the example of Fig. 3-1.

It is convenient to choose the following velocities with which to work:

V_1: The velocity of the left hand face (LHF) of the tail.

V_2: The velocity of the RHF of the tail and the LHF of the end cap 1, and the LHF of the stress rod.

V_3: The velocity of the LHF of the ceramic and the RHF of end cap 1.

V_4: The velocity of the RHF of the ceramic and the LHF of end cap 2.

V_5: The velocity of the RHF of end cap 2, the RHF of the stress rod, and the LHF of the head.

V_6: The velocity of the RHF (the driving face) of the head and the velocity of the case, assumed rigidly attached to the head.

It is now possible, with the aid of the points at which we have specified the velocity, to sketch the mechanical circuit shown in Fig. 3-3 using the impedance analogy.

To arrive at Fig. 3-3, each element (save the case) of the transducer has been divided into two halves, a right half and a left half. Thus, the notation is

M_{TL} = Mass of left half of tail,

M^{1}_{ER} = Mass of right half of end cap one,

M_{HL} = Mass of left half of head,

M_{CL} = Mass of left half of entire ceramic element,

M_{SRL} = Mass of left half of stress rod,

and so on.

Additionally, the reaction due to radiation is expressed as a force. The spring constants S_{ji} are obvious and since the ceramic is a force producing element, a force F is drawn as acting within it.

It should be noted that Fig. 3-3 is not a "real" representation of the transducer, basically because it shows (unrealistically) each element as consisting of rigid masses separated by springs, when in fact they are not, and are <u>not represented as such</u> when the final analysis is made. This figure is used only to devise the equivalent circuit of the transducer,

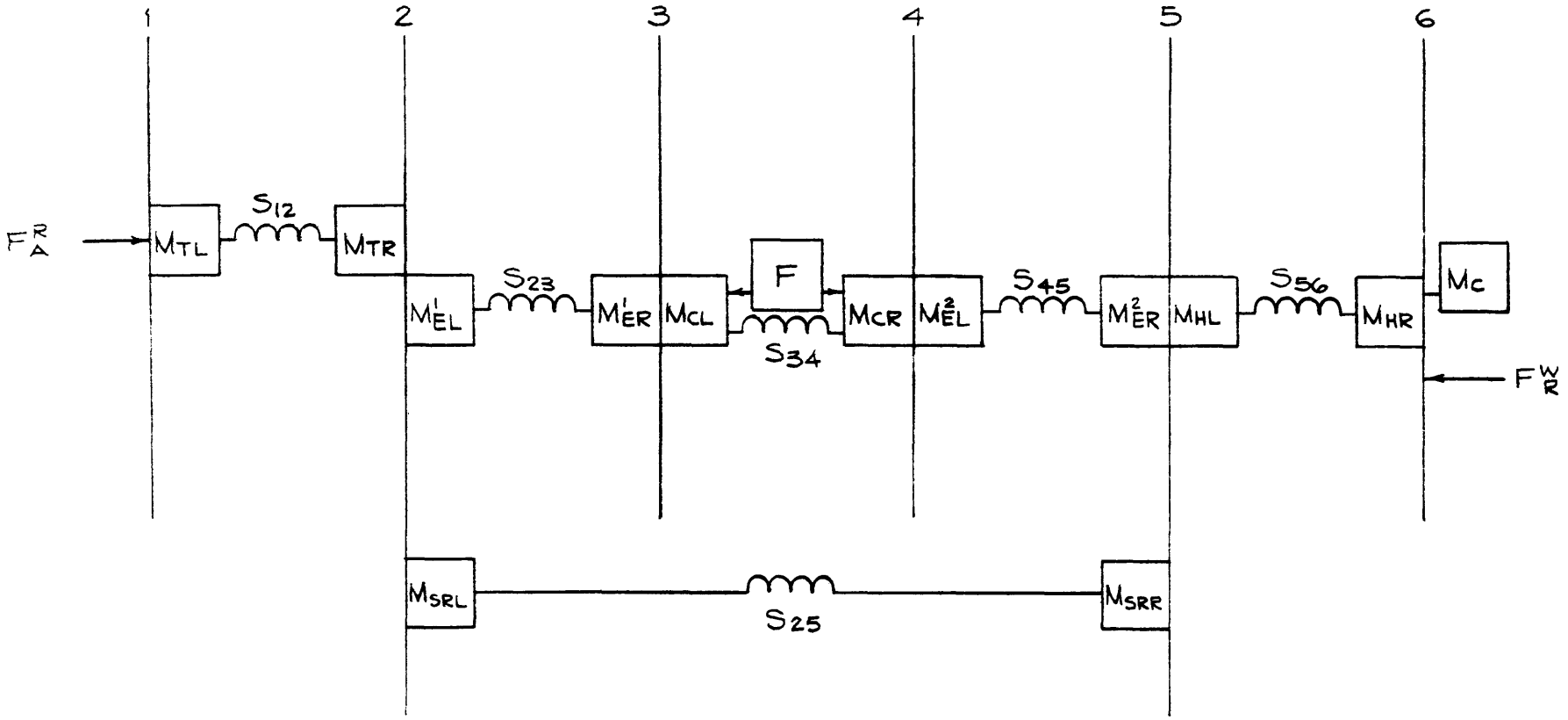


Fig. 3-3-MECHANICAL EQUIVALENT TO TRANSDUCER

TRACOR, INC. AUSTIN, TEXAS
DWG. 771-39
5-28-65 BMB ES

after which the artificial masses and springs are replaced by the real equivalent circuits of the compacted elements. The correct representations must be derived under the assumption that the elements are elastic bodies, not rigid masses and springs.

The derivation of the equivalent circuit now proceeds: At point 1, the force equation is

$$F_a = S_{12} (U_1 - U_2) + M_{TL} \ddot{U}_1 , \qquad (3\text{-}20)$$

where U_i, $i = 1,\ldots,6$ is the displacement from equilibrium position.

It is assumed that the displacement has a harmonic time dependence $e^{i\omega t}$, hence the velocity is given by

$$v = i\omega U, \text{ where } \frac{dU}{dt} = i\omega U$$

so that

$$F_a = \left(\frac{S_{12}}{i\omega}\right) (v_1 - v_2) + (i\omega M_{TL}) v_1 . \qquad (3\text{-}21)$$

Equation (3-11) can be written as

$$F_a = Z_{12} (v_1 - v_2) + Z_{11} v_1 \qquad (3\text{-}22)$$

$$Z_{11} \equiv i\omega M_{TL} \equiv Z_{TL}, \; Z_{12} \equiv S_{12}/i\omega ,$$

in direct analogy with an electrical circuit, i.e.

$$F \rightarrow \text{voltage}$$
$$v \rightarrow \text{current}$$

so that $F = Zv$, where Z is an impedance, either masslike (inductive) or springlike (capacitive).

The following equation holds at point 2,

$$0 = S_{12}\,(U_2 - U_1) + S_{23}\,(U_2 - U_3) + S_{25}\,(U_2 - U_5)$$

$$+ (M_{TR} + M_{EL}^{(L)} + M_{SRL})\,\ddot{U}_2 \qquad (3\text{-}23)$$

which in the same way as Eq. (3-12) becomes

$$0 = Z_{12}\,(v_2 - v_1) + Z_{23}\,(v_2 - v_3)$$

$$+ Z_{25}\,(v_2 - v_5) + Z_{22}\,v_2 \qquad (3\text{-}24)$$

where

$$Z_{22} \equiv Z_{TR} + Z_{EL}^{(1)} + Z_{SRL}\ .$$

From similar derivations, there are four more equations

$$-F = Z_{23}\,(v_3 - v_2) + Z_{33}\,v_3 + Z_{34}\,(v_3 - v_4)$$

$$Z_{33} = Z_{EL}^{(1)} + Z_{CL} \qquad (3\text{-}25)$$

$$F = Z_{45}\,(v_4 - v_5) + Z_{34}\,(v_4 - v_3) + Z_{44}\,v_4$$

$$Z_{44} = Z_{CR} + Z_{EL}^{(2)} \qquad (3\text{-}26)$$

$$0 = Z_{45}\,(v_5 - v_4) + Z_{56}\,(v_5 - v_6)$$

$$+ Z_{25}\,(v_5 - v_4) + Z_{55}\,v_5\ . \qquad (3\text{-}27)$$

$$Z_{55} = Z_{ER}^{(2)} + Z_{HL} + Z_{SRR}$$

$$-F_R^w = Z_{56}\,(v_6 - v_5) + Z_{66}\,v_6 \qquad (3\text{-}28)$$

$$Z_{66} = Z_{HR} + Z_{case}$$

The seven equations, Eqs. (3-22) to (3-28), together with an equation for F in terms of the current or voltage into the ceramic, completely define the equivalent circuit for the transducer. It is up to the analyst, by guess, hunch, inspiration or experience to come up with a circuit that fits the above equations. Examination will show that the circuit of Fig. 3-4 does in fact fit the equations.

3.5 AN IMPROVED TRANSDUCER MODEL

A much more refined approach to transducer analysis has been developed during the past few years at NEL, principally through the efforts of Martin, Carson, and others. In the new approach, the lumped parameter notion is abandoned in favor of treating each component of the transducer element, i.e., the head, tail, stress rod, case, ceramic, etc., as a continuous medium. Equations of elasticity then describe the motion of the inactive components and appropriate equations of state are used to describe the behavior of the active component (crystal or ceramic). Needless to say, this analysis is a much more realistic representation of the real transducer element. With this approach one can determine not only the "action" of the transducer, i.e. the variation of head velocity with input current or voltage and its bandwidth, but also the velocity and stress distributions of the various components. Such an approach is therefore extremely valuable from a design standpoint; and, with the availability of digital computers to perform rapid numerical analysis, the techniques can even be extended to tolerance studies. Such a study has extremely important implications in establishing manufacturing and testing standards.

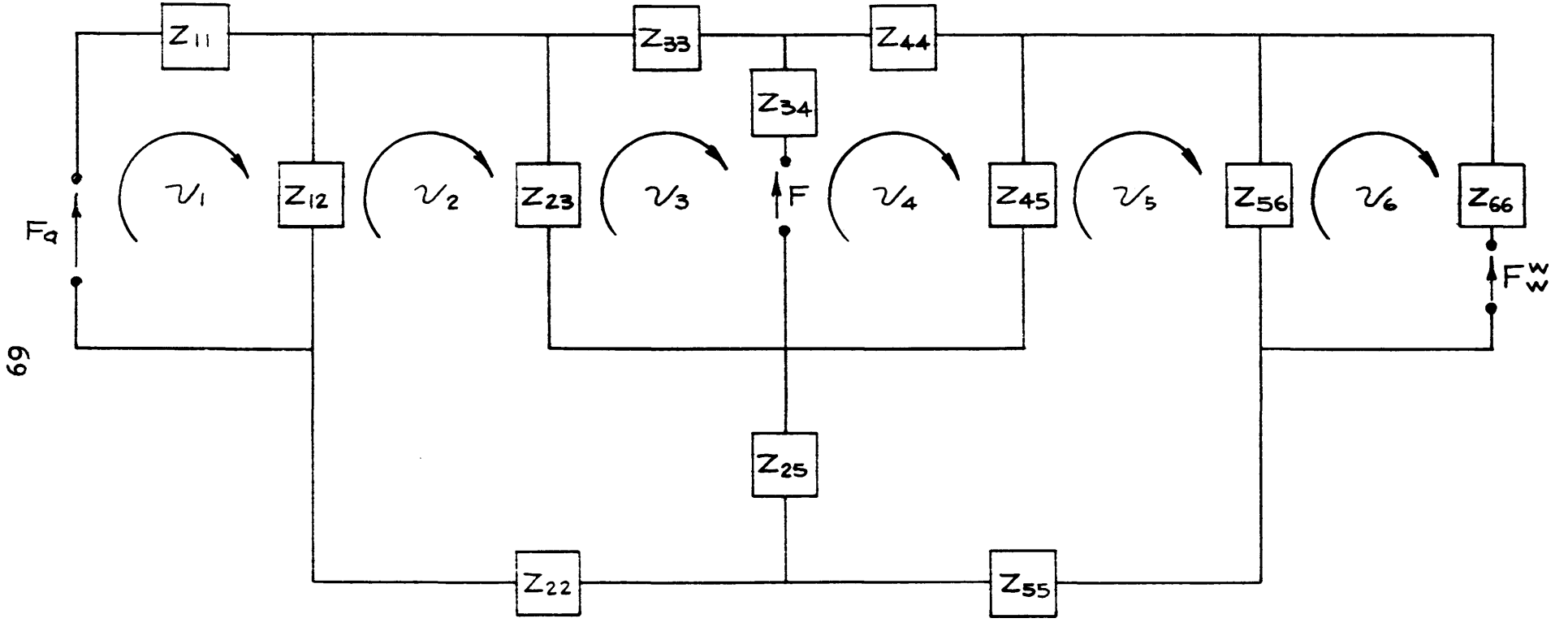


Fig.3-4-ELECTRICAL EQUIVALENT TO TRANSDUCER

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-40 5-28-65 BMB-ES

3.5.1 Inactive Components

The inactive components of the element (head, tail, stress rod, and case) are treated as elastic members. Equation (3-13) is the differential equation describing the motion of these components. The motion of these components in terms of the end forces and end velocities is given by Eqs. (3-14) and (3-15). The analysis of the combination of components is exactly the same as for the lumped parameter model in equations (3-22) to (3-28). The only difference in results for the lumped and distributed parameters is in the values for the various impedances. The functional relationship of forces and velocities is unchanged.

3.5.2 Active Components

In the active (piezoelectric, electrostrictive, or magnetostrictive) component, electric currents and voltage must be taken into account in addition to velocities, stresses, and strains. There is, nevertheless, some overlap between the active and inactive theory.

The velocity distribution in the active element is again found by solving the one-dimensional wave equation for the appropriate cross sectional area function and specifying boundary conditions at the ends of the element, just as for the inactive components. The strain is found by application of Hooke's law,

$$S_{xx} = \frac{1}{i\omega} \frac{\partial v}{\partial x} \quad ,$$

where S_{xx} is the normal strain component in the x-direction. However, the normal stress component in the x-direction, T_{xx}, is no longer simply related to S_{xx} by the Young's modulus, because of the peculiar nature of the material. The new relationship is of the form

$$T_{xx} = Y\, S_{xx} + T'_{xx}$$

where T'_{xx} is a function of the current or voltage to the material.

Similarly, the electrical equation is also modified by the nature of the material. One would expect the material between the foils on either side of the ceramic to give rise to some fixed capacitance C_o such that the circuit shown in Fig. 3-5 would apply, with

$$Z_o = \frac{1}{i\omega C_o}$$

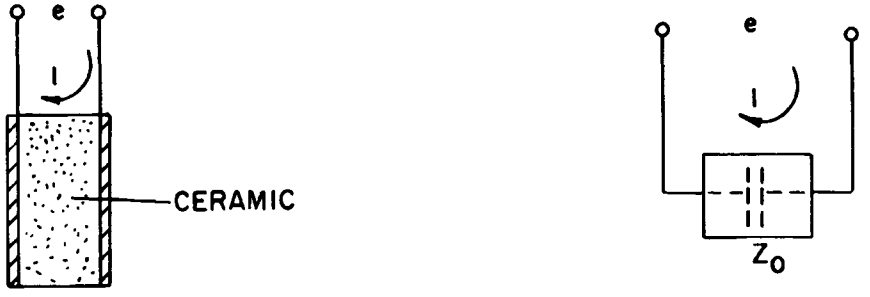


FIG. 3-5

This is not a complete description of its behavior because the element must absorb energy, hence there is a resistive component, and the mass of the ceramic vibrates, hence there is an inductive component.

The interactions of the mechanical and electrical quantities are described by the circuit of Fig. 3-6.

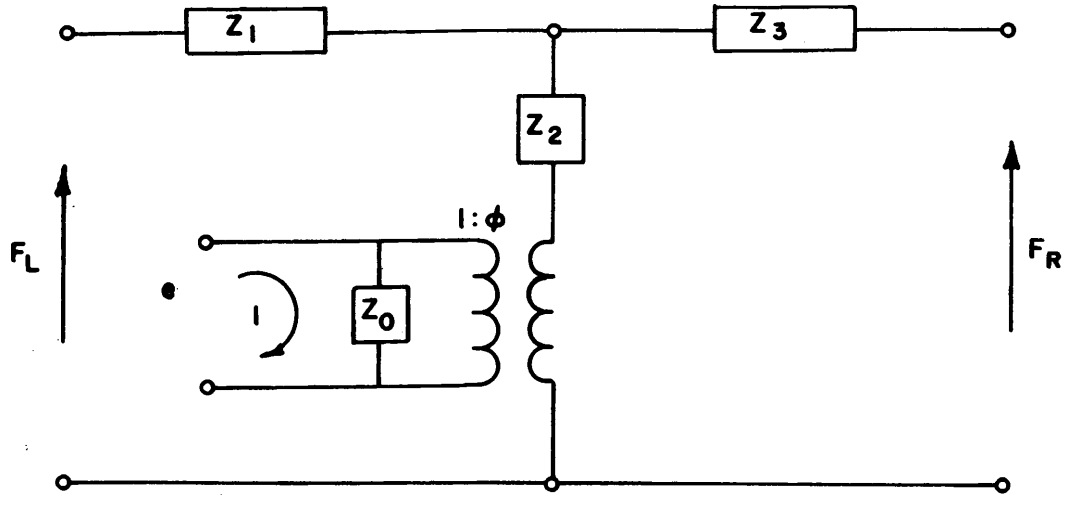


FIG. 3-6

The quantity φ, called the electro-mechanical transformation ratio, is a function of the descriptive parameters of the active material, and, among other things, makes the units compatible between the electrical and mechanical circuits.

Two plausibility arguments may be given for the circuit of Fig. 3-6. First, consider the case in which the applied voltage and current are zero, and the electrical leads a and b are, in effect, open. Then the electrical circuit simply becomes the impedance Z_o connected across the leads of the transformer. This may obviously be converted to a mechanical impedance and lumped with Z_2 to reduce the circuit to one similar in form to that in Fig. 3-1c for the inactive components.

Next, consider the case where the material is not allowed to vibrate ($v_L = v_R = 0$). Since the mechanical "current" in one coil of the transformer is zero, the electrical current in the other coil must also be zero, and the electrical circuit reduces to that of Fig. 3-5 as should be expected.

The equivalent circuit of the transducer element as derived by the distributed parameter analysis is represented in Fig. 3-4. This circuit is identical to the one derived by lumped parameter analysis except that much better estimates of the various impedances have been determined. In addition the velocity, displacement, and stress distribution in the element components can now be determined. This feature is, of course, not available in the lumped parameter analysis.

4. TRANSDUCER ARRAYS

In this section the characteristics of large transducer arrays are discussed. Beamforming techniques, shading, and typical array geometries are described. In addition the sonar environment, specifically ship self noise, is discussed along with a description of some studies directed toward quieting the sonar noise environment. A brief explanation of the effects of the interaction of the dome and transducer on sonar performance also is given.

4.1 ELEMENTARY ARRAYS

For transducers operating in the transmit mode it is desirable to direct the transmitted power into a small region in order to increase the range of the device, while in the receive mode it is desirable for the response of the transducer to be greater in one direction thereby reducing the level of the minimum detectable signal or increasing the signal-to-noise ratio. If one considers two harmonic, point sources of equal amplitude and phase lying on a straight line, then it is easy to show that the pressure P at great distances from the point array is given by

$$P = \rho \frac{\sin(\frac{\pi d}{\lambda} \sin \theta)}{\sin(\frac{\pi d}{2\lambda} \sin \theta)} , \qquad (4\text{-}1)$$

where θ is the angle between the field point and the axis of the array, ρ is the source strength, d is the separation of the sources, and λ is the wavelength of the transmitted energy. If one plots the pressure amplitude P(d) as a function of θ on a polar grid, curves like those shown in Fig. 4-1 are obtained. Hence it appears that the pressure distribution in the water can be altered by varying the separation d.

Now if the two points are replaced by a line source of

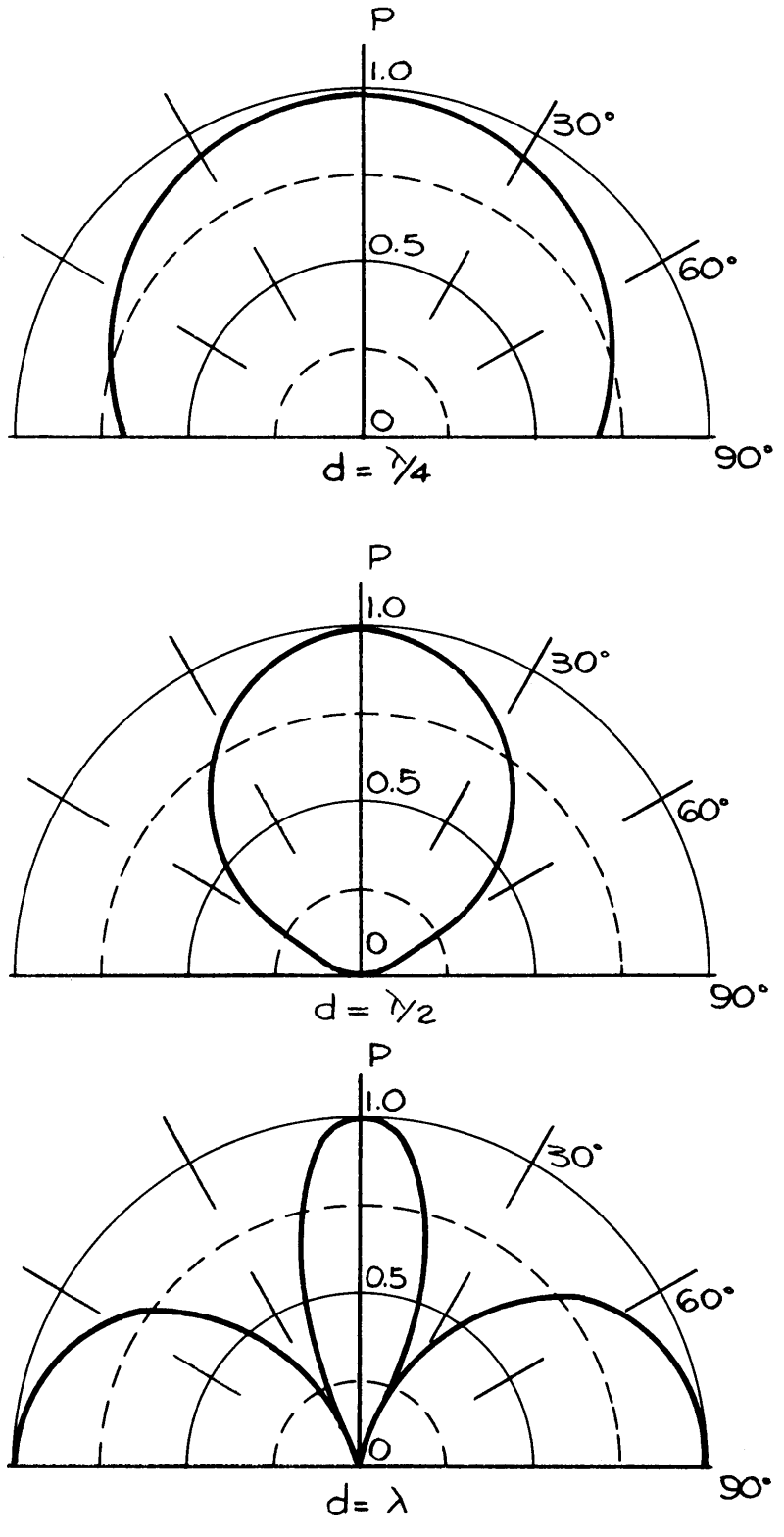


Fig.4-1-DIRECTIONAL CHARACTERISTICS OF TWO POINT SOURCES, "d" APART.

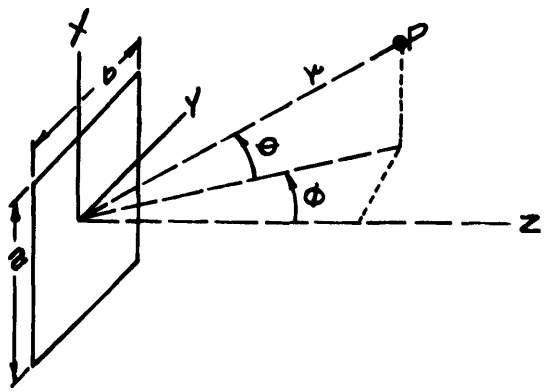
TRACOR, INC. | DWG. 771-26
AUSTIN, TEXAS | 5·27·65 BMB-ES

length ℓ each point of which has equal amplitude and phase, the pressure at a great distance is expressed as

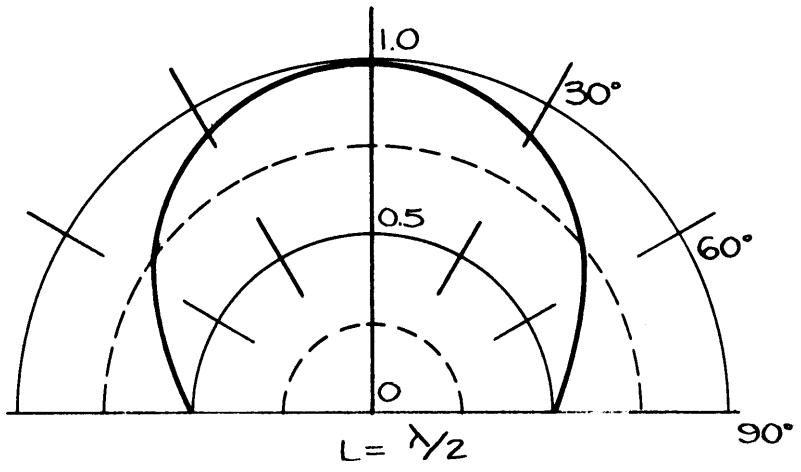
$$P = \int_{-\ell/2}^{\ell/2} \rho\, e^{-\frac{i2\pi x}{\lambda}\sin\theta}\, dx = \rho\ell \frac{\sin\left(\frac{k\ell\sin\theta}{2}\right)}{\frac{k\ell\sin\theta}{2}} \qquad (4\text{-}2)$$

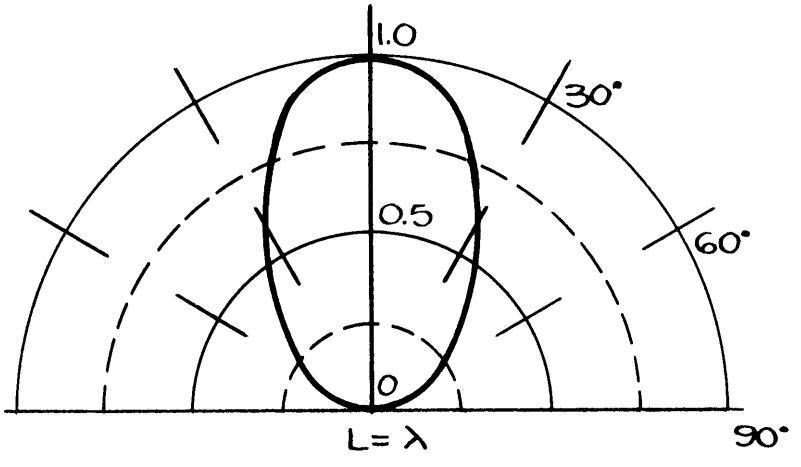
where $k = \frac{2\pi}{\lambda}$ is the wave number, and ρ is the source strength of a point in the line. A few typical pressure patterns obtained from this relation are shown in Fig. 4-2. Here it is evident that the radiation pattern can be made directive, i.e., the main lobe narrowed, by making the source length larger. However, the appearance of side lobes should be noted.

Many transducers have a two-dimensional surface rather than a line configuration; i.e., the transducer is an area transmitter or receiver. In this case the radiation pattern or receive pattern is three-dimensional. For example, a circular disc radiator has a transmit or receive pattern that is a solid of revolution, such as shown in Fig. 4-3. A rectangular radiator, as shown below,



has a pressure amplitude distribution given by





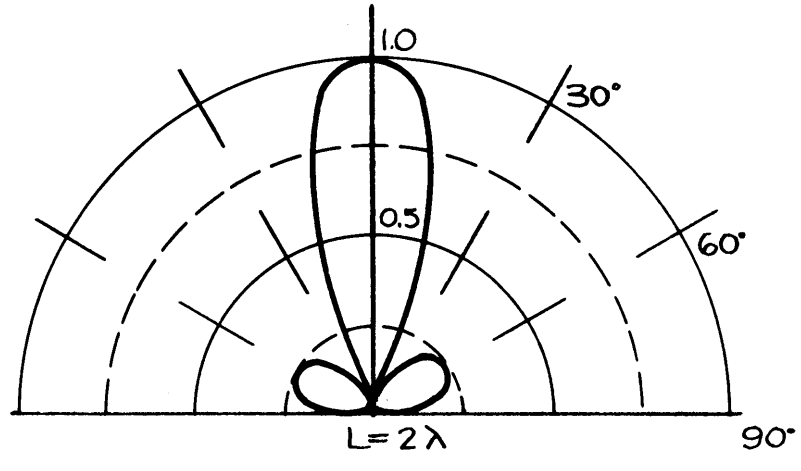


Fig 4-2-DIRECTIONAL CHARACTERISTICS OF A LINE SOURCE OF LENGTH "L."

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-27 | 5-27-65 BMB-ES

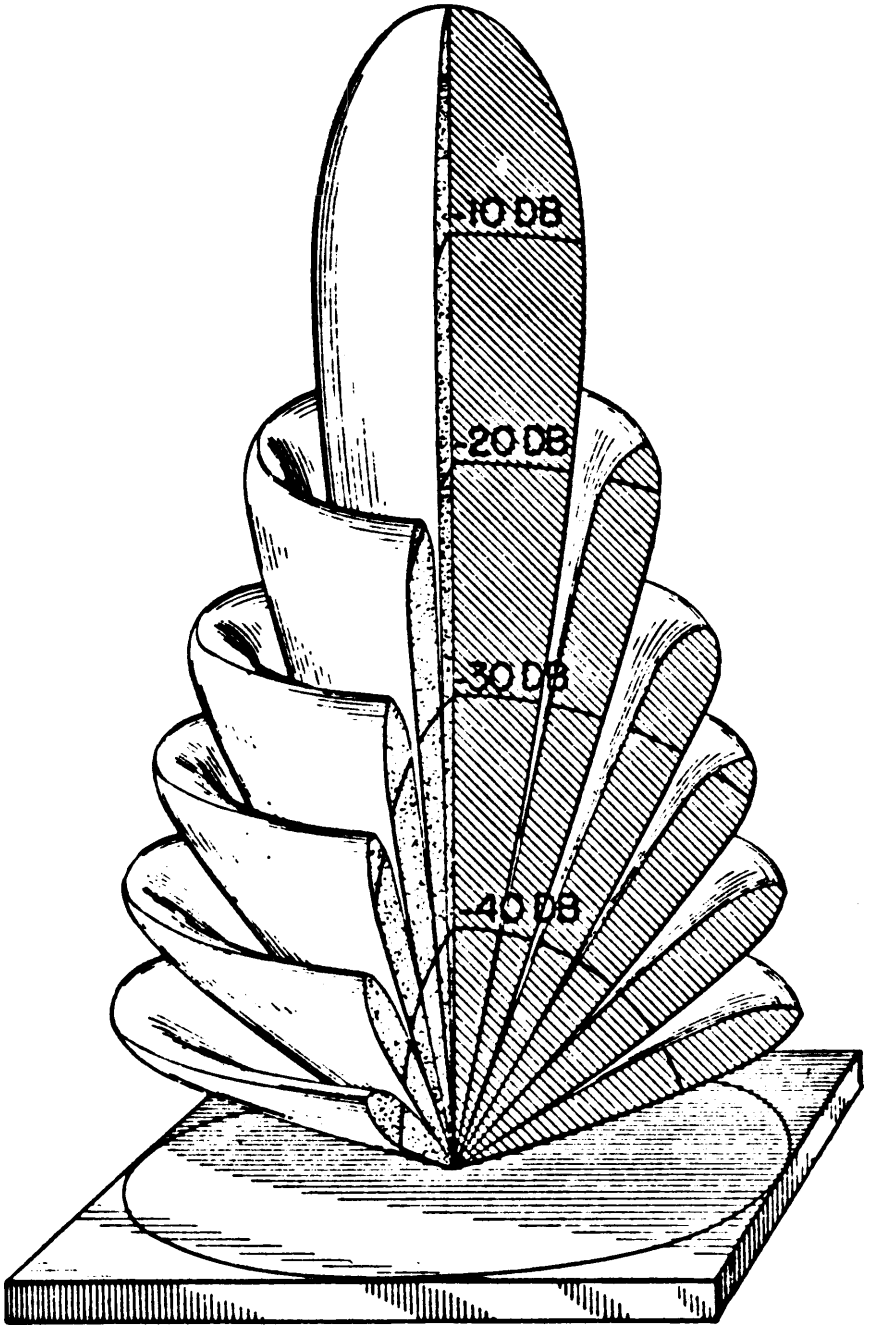


FIG. 4-3 – BEAM PATTERN OF A CIRCULAR DISK SOURCE

TRACOR, INC DWG A771-9
AUSTIN, TEXAS 12/23/64 SFF/AA

$$P(r,\theta,\varphi) = ab \left[\frac{\sin(\frac{ka\sin\theta}{2})}{\frac{ka\sin\theta}{2}}\right] \left[\frac{\sin\frac{kb\sin\varphi}{2}}{\frac{kb\sin\varphi}{2}}\right]$$

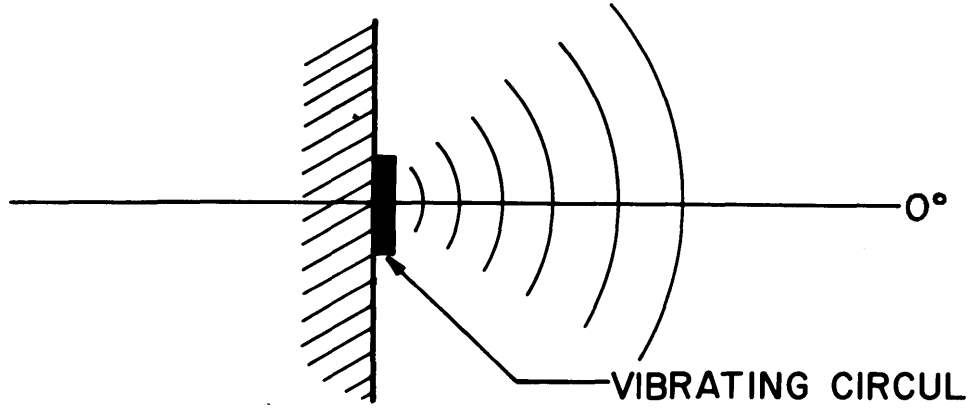
assuming each point in the radiator is vibrating with unity amplitude and in the same phase. Several ways of plotting beam patterns are employed; the polar and rectangular forms are shown in Fig. 4-4. Some standard nomenclature of beam patterns is shown in Fig. 4-5.

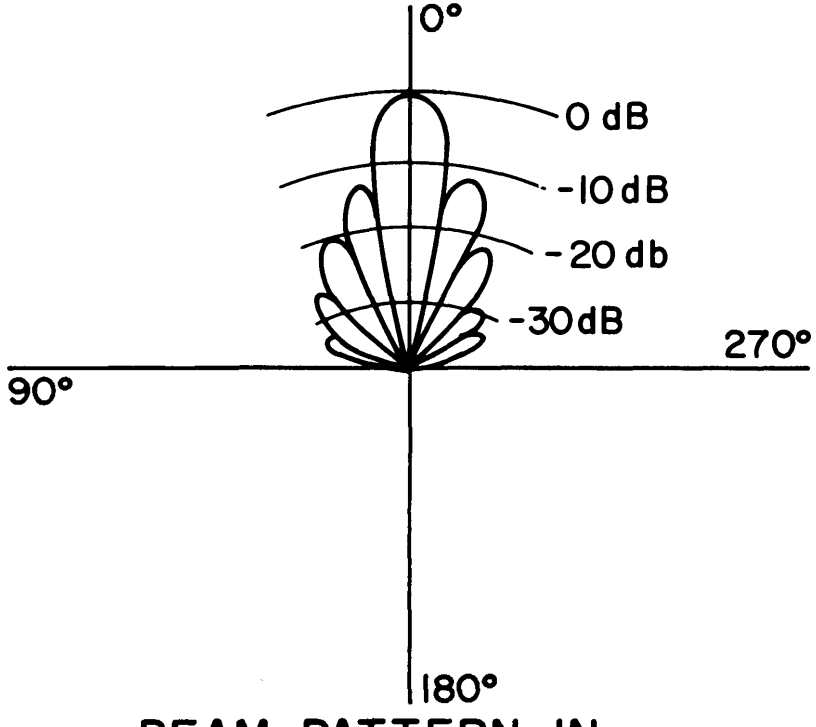
It is convenient to talk about the directional property of a transducer in terms of a quantity called the directivity index. The directivity index has been defined in a previous section of the report but will be repeated here. The directivity index, D.I., is defined for the transmit case as

$$D.I. = 10 \log \frac{\text{peak intensity of the radiation pattern}}{\text{avg. intensity of the radiation pattern}} \quad (4\text{-}3)$$

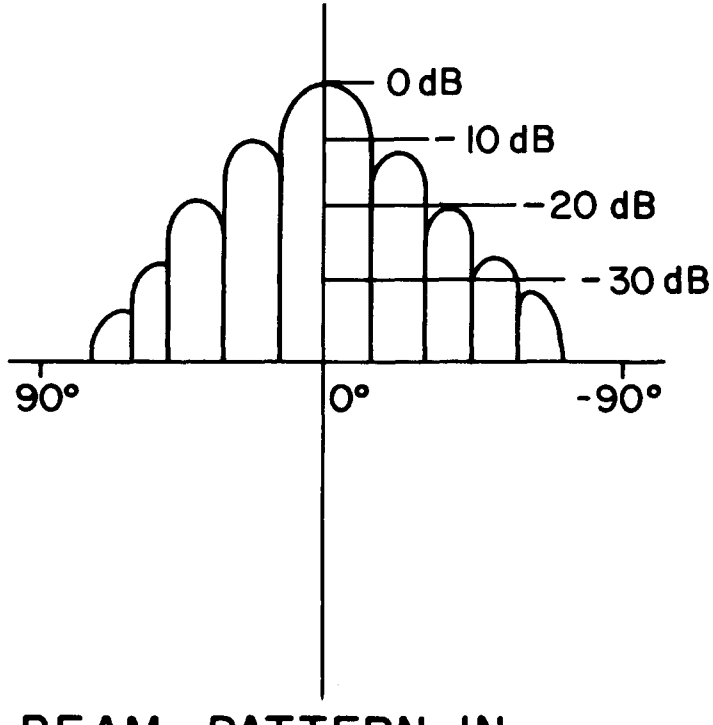
This definition is applicable to one- or two-dimensional radiators. For the receive case, D.I. is defined as the negative of the above. This latter definition is possible since there is a one-to-one correspondence between the pressure incident on a transducer element and the output voltage of that element. (In the receive case a pattern would be obtained by measuring the output voltage of the transducer elements.)

The trend in sonar development has been toward designing sonar equipments with increased range. Since the power a transducer can put into the water is limited by cavitation, an increase in range is possible only by going to lower frequencies (longer wavelengths) for which attenuation losses are decreased, or larger sized transducers, or a combination of both. It is apparent from the radiation pattern equation for a line source that as the





BEAM PATTERN IN POLAR COORDINATES



BEAM PATTERN IN RECTANGULAR COORDINATES

FIG. 4-4

TRACOR, INC
AUSTIN, TEXAS
DWG A771-10
12/23/64 SFF/AA

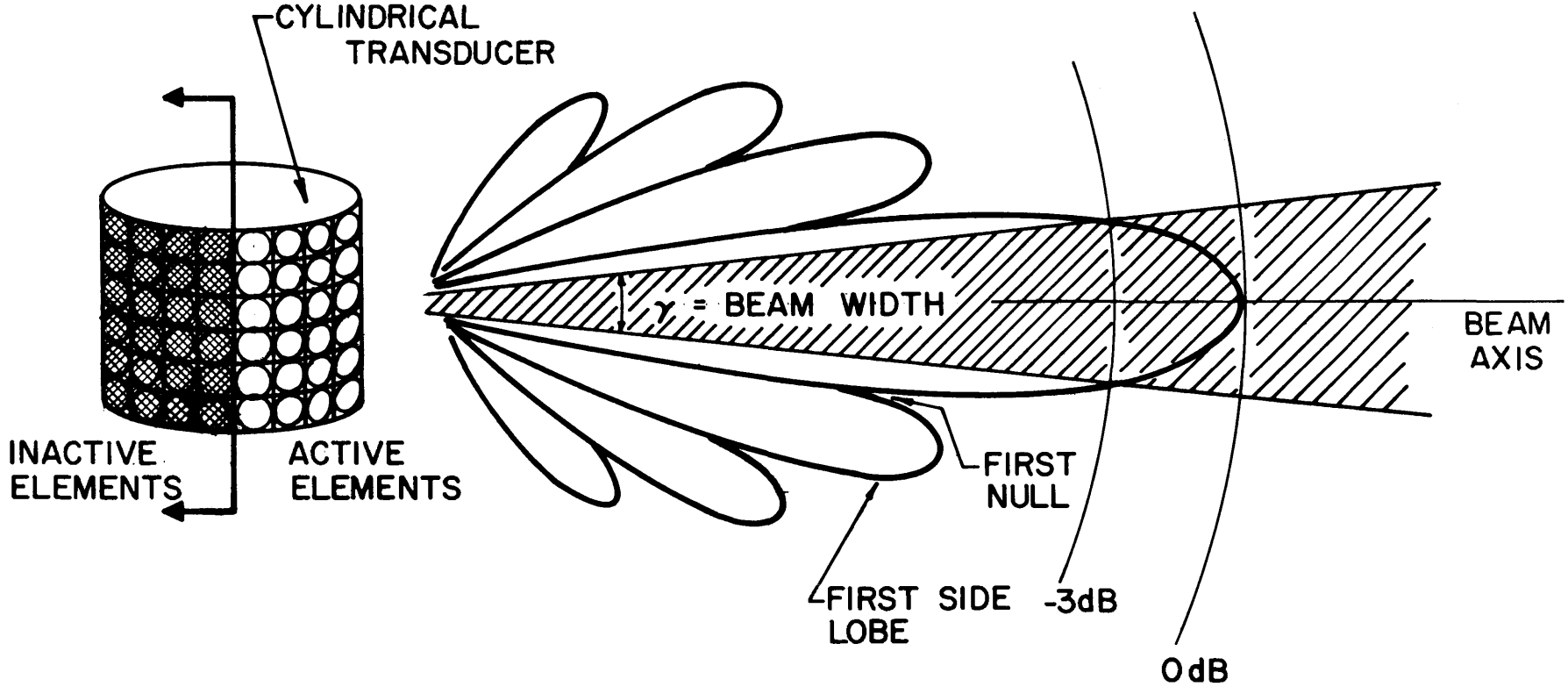


FIG. 4-5 –BEAM PATTERN NOMENCLATURE

wavelength is increased the physical dimensions of the source must increase in order to maintain the same radiation pattern. As a result sonar transducers have grown from relatively small devices to equipments weighing many tons and having dimensions of several feet.

Early sonar transducers were of the searchlight type; that is, directional coverage was obtained by mechanically steering the transducer, usually composed of a small number of elements. However, as technology improved and the size of the required transducer increased, the transducer evolved as a large set of elements. For a fixed transducer dimension there are many reasons for preferring several small elements to one large element. Among these are

1. The problem of directional coverage can be shifted from mechanical steering to electrical scanning.

2. The loss of a few elements in an array of many elements does not seriously affect transducer performance, whereas the loss of the element in a single element transducer is catastrophic.

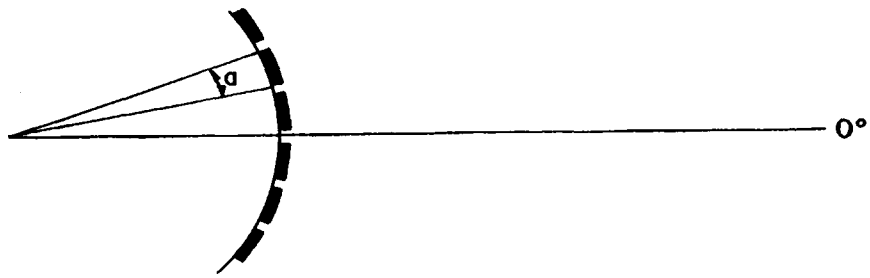
3. The transducer directivity pattern can be improved through a reduction of sidelobe level by shading the elements.

4. The obvious physical limitations in trying to build an element which in modern applications has radiating face dimensions of several feet are avoided.

The use of several small elements rather than one large one does introduce problems, however. These will be discussed elsewhere.

4.2 BEAMFORMING

Consider now an array of small elements, say in the shape of an arc of a circle. (Common element sizes range from $3\lambda/8$ to $\lambda/2$. The spacing between elements is quite small, the edges of the elements being much less than a wavelength apart.)



Each of these elements has a radiation pattern similar to those already described. If these elements oscillate in phase and with equal amplitude, the array will have a radiation pattern which is approximated by the relation

$$p(\theta) = \frac{1}{2m+1}\left| \sum_{k=-m}^{m} \cos\left[\frac{2\pi R}{\lambda}\cos(\theta + k\alpha)\right] \frac{\sin[\frac{\pi d}{\lambda}\sin(\theta + k\alpha)]}{\frac{\pi d}{\lambda}\sin(\theta + k\alpha)} \right.$$

$$\left. + i \sin\left[\frac{2\pi R}{\lambda}\cos(\theta + k\alpha)\right] \frac{\sin[\frac{\pi d}{2}\sin(\theta + k\alpha)]}{\frac{\pi d}{\lambda}\sin(\theta + k\alpha)} \right| \qquad (4\text{-}4)$$

where $P(\theta)$ is the ratio of the pressure amplitude at the angle θ to the pressure amplitude at $\theta = 0^o$, R is the radius of curvature of the array, α is the angular spacing of the elements, d is the element length and $2m+1$ is the number of elements. It is assumed that one element is positioned in the direction $\theta = 0^o$ and the remaining $2m$ elements are positioned symmetrically about $\theta = 0^o$. A sample graph of a 60^o arc source having radius of 4λ is shown in Fig. 4-6. This array has a "projected" length of 2λ. A line array 2λ long has a pattern as shown on the curves in Fig. 4-2. It is clear from these pictures that a linear array is more desirable, in terms of directivity, than is a curved array having the same projected length.

It is possible to reduce the arc source to an equivalent line source by a simple phase delay of the elements and thereby improve the radiation pattern. If the ith element is advanced

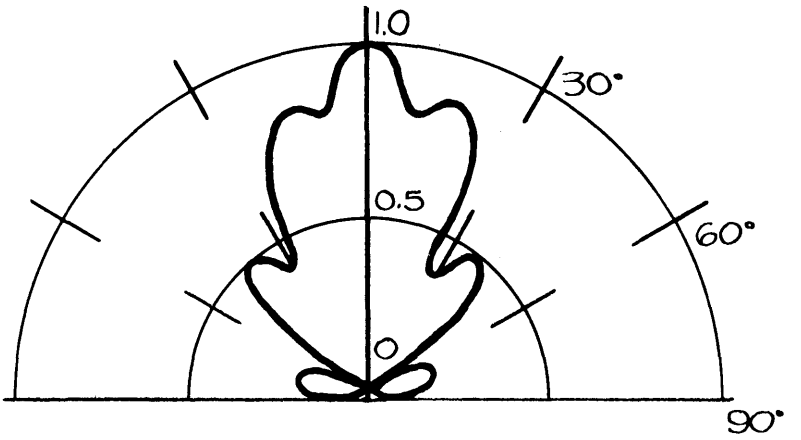
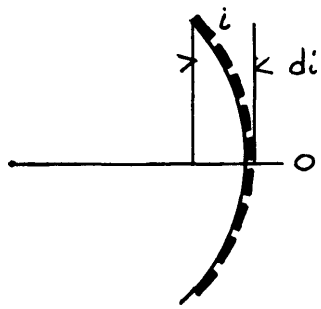


Fig. 4-6 DIRECTIONAL CHARACTERISTICS OF A 60° ARC OF RADIUS 4 λ.

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-28 5-27-65 BMB-ES

in phase an amount $2\pi d_i/\lambda$ then it will be in phase with the zeroth element. If all elements are adjusted in this fashion, the array will produce a far-field pattern in a manner equivalent to a straight line array. This phasing process is an elementary operation in beamforming and is typical for modern sonar equipments.

A cylindrical array of elements in which the elements are arranged in horizontal rows and in columns (staves) is a useful transducer form.

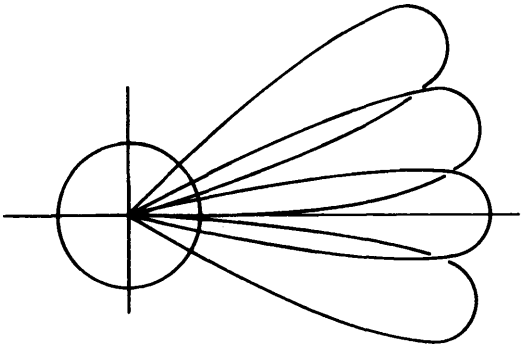


Suppose we want to receive or transmit in the 0^o direction and in the horizontal plane (The beamforming problem is the same for transmit and receive.) In order to transmit, a certain number n of elements are phased to represent a planar source. In practice this usually is about one-third of the total for a typical narrow beam. To transmit in some other direction, say α, n elements centered about the direction α are driven with proper phasing. It can be seen that if the adjusting is done electronically the beam can be steered very rapidly by indexing one element at a time so that the array effectively has 360^o coverage with the same directivity as an equivalent straight line array. The above discussion has been limited to patterns in the horizontal plane (zero tilt angle). Many arrays can also be phased to transmit or receive at other tilt angles for use in convergence zone or bottom bounce modes. In the latter cases, the elements are phased to a plane perpendicular to the direction (tilt angle and bearing

angle) of transmission or reception.

The elements in a cylindrical array are in the form of a cylindrical matrix the rows of which are called layers and the columns called staves. Beamforming in the receive case is done by summing the signals of the elements in each stave, properly phased; then the stave signals are summed with appropriate phasing. The sum of the element outputs is a voltage representing the response of the array, in the direction it is looking, to the sound pressure field incident on the array.

In addition to scanned or steered beams, fixed beams may be employed. In this technique the outputs of the groups of elements are used to form beams looking or transmitting in fixed directions as diagrammed below.



It should be noted that a distinction is also made between analog and digital beamforming. In the analog case the case the output of each element is treated in its "real" form while in the digital case the output of each element is sampled and the beamforming operation is carried out digitally.

In the receive case the beamforming process represents a spatial correlation which discriminates against incoherent noise incident on the transducer, i.e., beamforming increases the signal-to-noise ratio. This processing gain or signal-to-noise improve-

ment is of the order of 25 dB in common sonar systems.

4.3 TRANSDUCER SHAPES

Perhaps the impression to this point, is that all transducers are cylindrical. Actually, the application of the sonar equipment determines to a large extent its shape. Although cylindrical shapes are prevalent, one also encounters spherical, planar, line, and conformal arrays. Surface-ship-mounted sonars are generally cylindrical. This is an obvious choice because for the surface ship the sonar must have rapid scan and complete, or near complete, 360^{o} coverage. Until recently the vertical coverage was limited to little more than the region above the thermal layer, so that the cylindrical shape was an obvious choice. The cylindrical array can be phased in the vertical direction to accommodate depressed beams. Depressed beams are not a particular problem for the cylindrical array because only a few discrete tilt angles must be handled. Such an array would not be efficient if continuous coverage in the vertical plane were required, because the vertical patterns suffer somewhat from "end fire" problems.

Spherical arrays are employed where continuous coverage in both horizontal and vertical planes is required as, for example, in submarines. Beamforming is accomplished in a manner similar to that done with cylindrical arrays.

Conformal arrays, which follow the contours of a ship, are usually employed only for the receive mode. Conformal arrays can provide extended apertures and high angular resolution. One problem with conformal arrays is that the effective array length is a function of the "look" angle, hence the directivity is a variable.

Line arrays are generally towed and are passive. These arrays are highly directive broadside but again their directivity depends on the look angle. Small steerable line arrays are also employed in submarine systems.

Planar arrays are used both passively and actively and can be steered moderately. Their coverage is limited in end fire as is the case with conformal and line arrays.

4.4 SHADING

It was noted in the radiation patterns for a straight line source that as the ratio d/λ is increased side lobes begin to appear. Large side lobes are clearly not desirable since they can lead to target beaming error, i.e., an ambiguity concerning the direction of maximum response of the array. What one would like to have is a narrow main lobe with no side lobes at all. This pattern cannot be obtained. However, through a technique known as shading it is possible to reduce the side lobes for a given array. Shading is accomplished by varying the amplitude of the signal supplied to the elements across the array. The shading factors applied to the signal amplitudes may follow, for example, a cosine or cosine squared law. A newer method known as the Dolph-Tschebyscheff method provides, relative to other shading procedures, the narrowest possible major lobe for a given side lobe reduction. In this method the signal amplitude distribution applied to the elements corresponds to a Tschebyscheff polynomial. The disadvantages of shading are that the power radiated into the water is reduced and that any type of shading is accompanied by some broadening of the major lobe.

4.5 VELOCITY CONTROL

The specifications often applied to a transducer include maximum power output for any particular mode of operation and receive and transmit directivity. It has already been stated that cavitation limits the power density. It is also true that the far-field radiation pattern depends on the velocity distribution of the transmitting device. To this point in the discussion it has been assumed that ideal (uniform) velocity distributions exist in transducer arrays. If, however, one blithely takes a group of transducer elements and builds an array hoping

to obtain some particular pattern, one is apt to be surprised, because of element interaction. Because each element produces acoustic pressure at each other element in the array, the elements produce forces on each other. Consequently, the load that any element "sees" changes depending on that element's position in the array. In extreme cases, some elements may become sinks rather than sources; i.e., they absorb acoustic energy. This behavior sometimes causes damage to power amplifiers in addition to modifying the desired velocity distribution of the array.

The total reaction force on the ith element in an array is (neglecting dome effects)

$$F^R_i = Z^R_i \, v_i + F_s' \tag{4-5}$$

where F^R_i is the total radiation force on the ith element, Z^R_i is the free field radiation impedance of the ith element (the free field radiation force is $Z^R_i \, v_i$ where v_i is the head velocity of the ith transducer element), and F_s' is the sum of the radiation forces exerted on the ith element by all other elements in the array. By analogy to the self impedance relation it is convenient to define the force f_i produced on the ith transducer by the jth one as

$$f_i = Z_{ij} \, v_j \, , \tag{4-6}$$

where $Z_{ij} (=Z_{ji})$ is the mutual impedance between transducer elements. Then

$$F^R_i = Z^R_i \, v_i + \sum_{j \neq i} Z_{ij} \, v_j \, . \tag{4-7}$$

If $Z_{ii} = Z^R_i$, then

$$F^R_i = v_i \left[\sum_j Z_{ij} \frac{v_j}{v_i}\right] \qquad (4\text{-}8)$$

From this relation it is clear that the radiation impedance at the ith transducer is dependent on the velocities of all the transducers in the array. Hence it is possible that, given a voltage, one has no idea what the head velocity of the transducer will be; or, given the head velocity, one does not know the required voltage, if the element head velocity is dependent on the radiation loading. The upshot of the matter is that the design must be conducted on an array level rather than an element level.

Ideally one would like either voltage velocity

$$V_i = f(v_i) \qquad (4\text{-}9)$$

or current velocity control

$$I_i = g(v_i) \qquad (4\text{-}10)$$

In either case the voltage or current uniquely defines the head velocity, or vice versa. From an analytical standpoint, the presence of velocity control enables one to analyze an array one element at a time, whereas in the absence of velocity control the transducer must be analyzed as a whole.

4.6 SONAR ENVIRONMENT

Mounted within the sonar dome on a surface ship, the transducer finds itself in a rather noisy environment which can severely limit its performance. Aside from ambient sea noise, the sonar platform itself generates a variety of noises ranging from single frequency or very narrow band machinery noise to broad band flow noise and cavitation noise. Clearly, the sonar per-

formance depends on the level and coherence of the ambient noise field, and an effective noise-quieting program must consider both facets of the ambient noise characteristics.

The paths by which noise is transmitted to the dome-transducer environment will be considered first. The noise is either machinery-associated or water-associated; i.e., the noise is a direct result of vibrating machinery or of phenomena associated with movement in the water. The paths by which the noise reaches the transducer area are either (a) directly mechanical, that is, through the ship structure and thence through the transducer mountings or (b) water-borne, that is, propagated from the source through the water and dome, thence to the transducer. It is possible, of course, for machinery generated noise to find its way to the transducer via a water path, viz. a piece of vibrating machinery can drive the hull which radiates into the water. The energy is subsequently transmitted, in the water, through the dome to the transducer.

Typical sources of ship self noise are

1. Machinery
2. Bubble Sweepdown
3. Surface Roughness
4. Flow Boundary Layer
5. Screws

A common technique of measuring noise level is to take the mean noise level of the forward 210^{o} of the self-noise directivity pattern. This measure of noise varies typically with ship speed as shown in Fig. 4-7. Generally, for speeds below 10 kts, the noise is produced by machinery. This noise tends to be in spectrally discrete frequencies or narrow bands. From 10 kts to 20 kts the noise results from flow around the hull and dome boundaries. At speeds above 20 kts screw noise and noise resulting from local cavitation, caused by surface roughness, predominates. This noise is broadband.

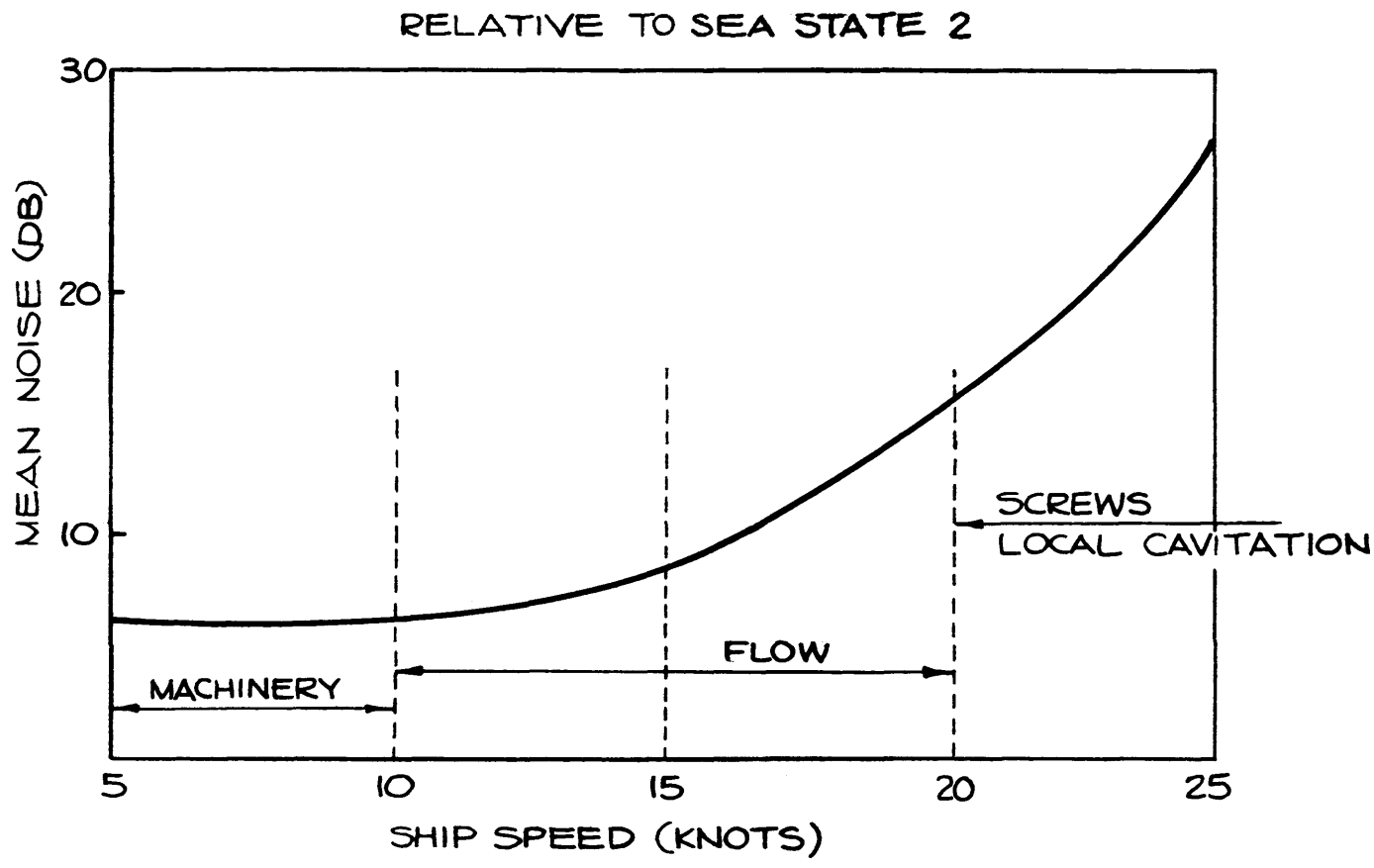


Fig 4-7-SELF NOISE BEHAVIOR

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-31 | 5-27-65 BMB-ES

It is difficult to describe a "typical" self-noise directivity pattern, because these patterns vary widely depending on ship's speed, dome-baffle configuration, surface condition of the dome, sea state, etc. However, an example is shown in Fig. 4-8. There has been, of course, a great deal of effort directed toward quieting the transducer environment. A typical quieting program consists of

1. Dome relocation: moving the sonar equipment away from noisy machinery.
2. Dome isolation: providing vibration damping between the dome and hull.
3. Transducer isolation: providing vibration damping between the transducer and the hull.
4. Hull damping: mounting vibration dampers on the hull.
5. Hull coatings: coating the hull with damping materials.
6. Dome damping: providing the dome with vibration dampers and coatings.
7. Baffles: shielding the transducer from screw noise.
8. Smoothing: reducing surface roughness on the hull and dome in an effort to reduce local cavitation.
9. Machinery isolation: providing damping to eliminate transference of machinery vibrations to the hull and structure.

These efforts have improved the problem somewhat, particularly for lower speeds. However, above 20 kts quieting programs have not produced significant results. In this region it may be recalled that ship's screws and local cavitation are the main culprits in producing noise in the sonar environment. Of course, smoothing the hull and dome help reduce local cavitation, but not a great deal has been accomplished in reducing screw noise.

The problem of the interaction of the sonar dome-transducer -baffle with screw noise has received considerable attention during

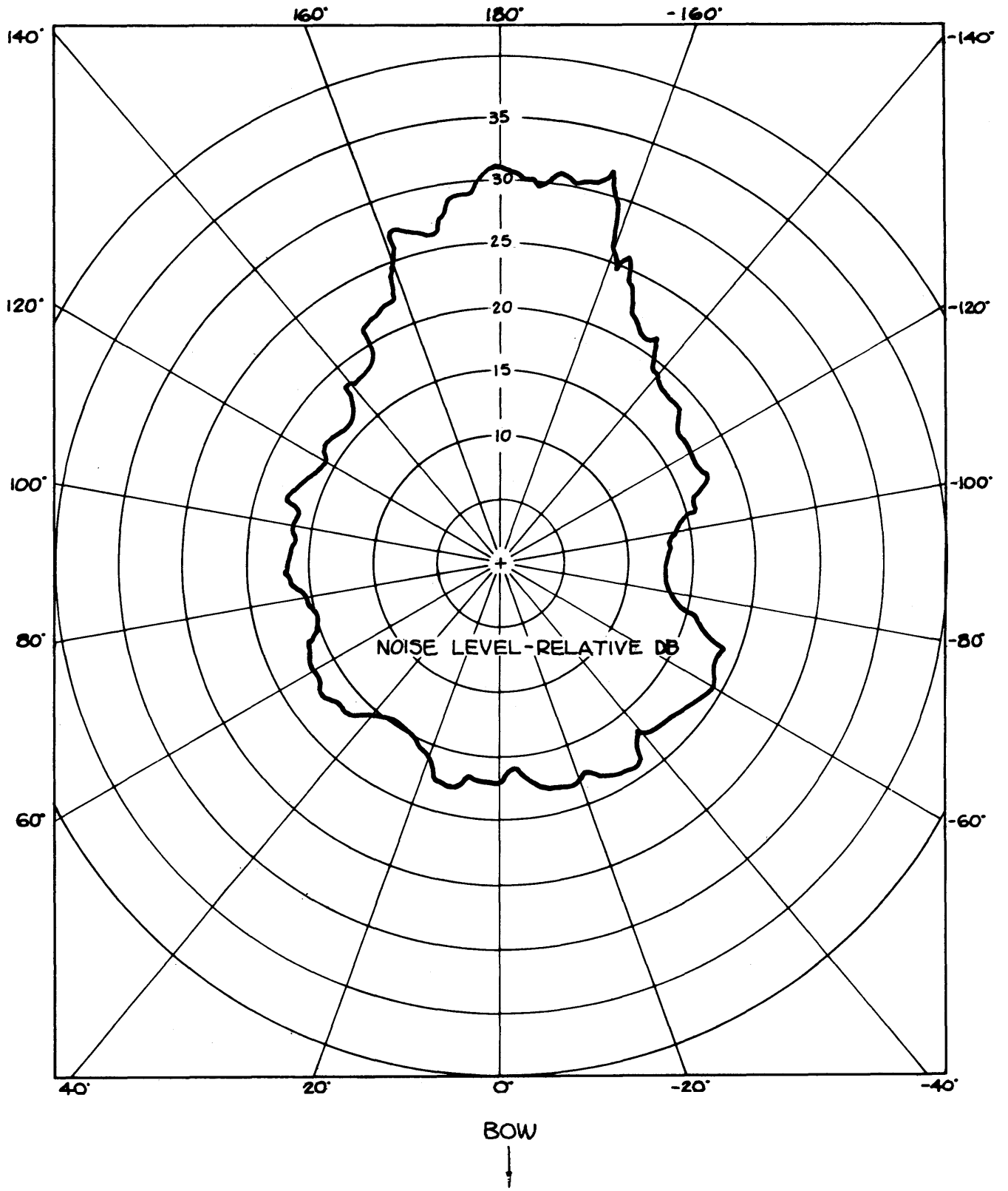


Fig. 4-8-TYPICAL SELF NOISE DIRECTIVITY PATTERN

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-32 5-28-65 BMB-ES

the past several years. The results of this study, while not designed to correct the problem, have furnished insight into the mechanism of dome-transducer-baffle interaction. In addition to the analytical studies conducted, significant experimental work in this area has been conducted by Dr. C. J. Burbank at NEL. In each of the studies an effort has been made to isolate the individual effects of the various components of the dome environment, for example the dome skin, supporting structure, and the baffle. In this way one can better understand how to take remedial steps to correct the ill effects of screw noise.

Among the results obtained in the experimental and analytical studies, it was found that the dome skin has little effect on self-noise. This conclusion was reached experimentally by measuring self-noise directivity patterns with the dome skin in place and with the skin removed. The conclusion was reached analytically by computing the sound pressure field resulting from the transmission of a prescribed field through plates and shells. It was found that the prescribed pressure field was basically unaltered.

The study of the interaction of the transducer and baffle with screw noise indicated that the baffle was in part responsible for the large spokes found in self-noise directivity patterns. The mechanism of this interaction involves the diffraction of screw noise around the edges of the baffle and transmission of the noise through the baffle (The transmitted noise energy is markedly attenuated by the absorbent coatings put on the baffle). The mathematical analysis involved in such a study typically begins by choosing a model which defines the geometry of the problem, in this case a cylindrical transducer and baffle. A solution is then obtained for the differential equation which describes the phenomenon under consideration, forcing the solution to satisfy boundary conditions specified on the model. In this case, a solution to the wave equation was obtained for boundary conditions specified on the

transducer and baffle assuming an externally incident pressure field (screw noise). The solution was then programmed for numerical evaluation on a digital computer, and results were obtained for several operational sonar equipments. A comparison of numerical results and experimental results obtained by Dr. Burbank for the interaction of the baffle and transducer with screw noise is shown in Fig. 4-9.

Experimental results also indicated that the dome supporting structure could cause spokes in self-noise directivity patterns. A typical supporting structure consists of a lattice structure of small-diameter bars. This periodic array of small scatterers acts like a diffraction grating and scatters incident energy in coherent wavefronts which produce a relatively high response on the sonar display. The dome supporting structure was studied mathematically using a model consisting of a periodic array of small rods. The scattered field of this array was computed for a prescribed incident field (screw noise). These results were combined with the results of the baffle-transducer interaction to achieve an effective dome-transducer-baffle interaction with screw noise. These latter results are compared with experimental results in Fig. 4-10.

During the past year a program of study has been conducted to determine the interaction of the dome and transducer during transmission. Here the problem has many features in common with previous studies; however, the increase in energy in the transmit mode can result in a significant excitation of the dome, producing an alteration in both the near- and far-field energy distributions. This study is not yet complete. Preliminary results, however, indicate that the levels of the side lobes in the far-field beam patterns are significantly increased, compared to the beam pattern for a bare transducer. The near-field pressure is also perturbed due to reflections from the dome, the pressure distributions on the transducer face becoming highly nonuniform, compared

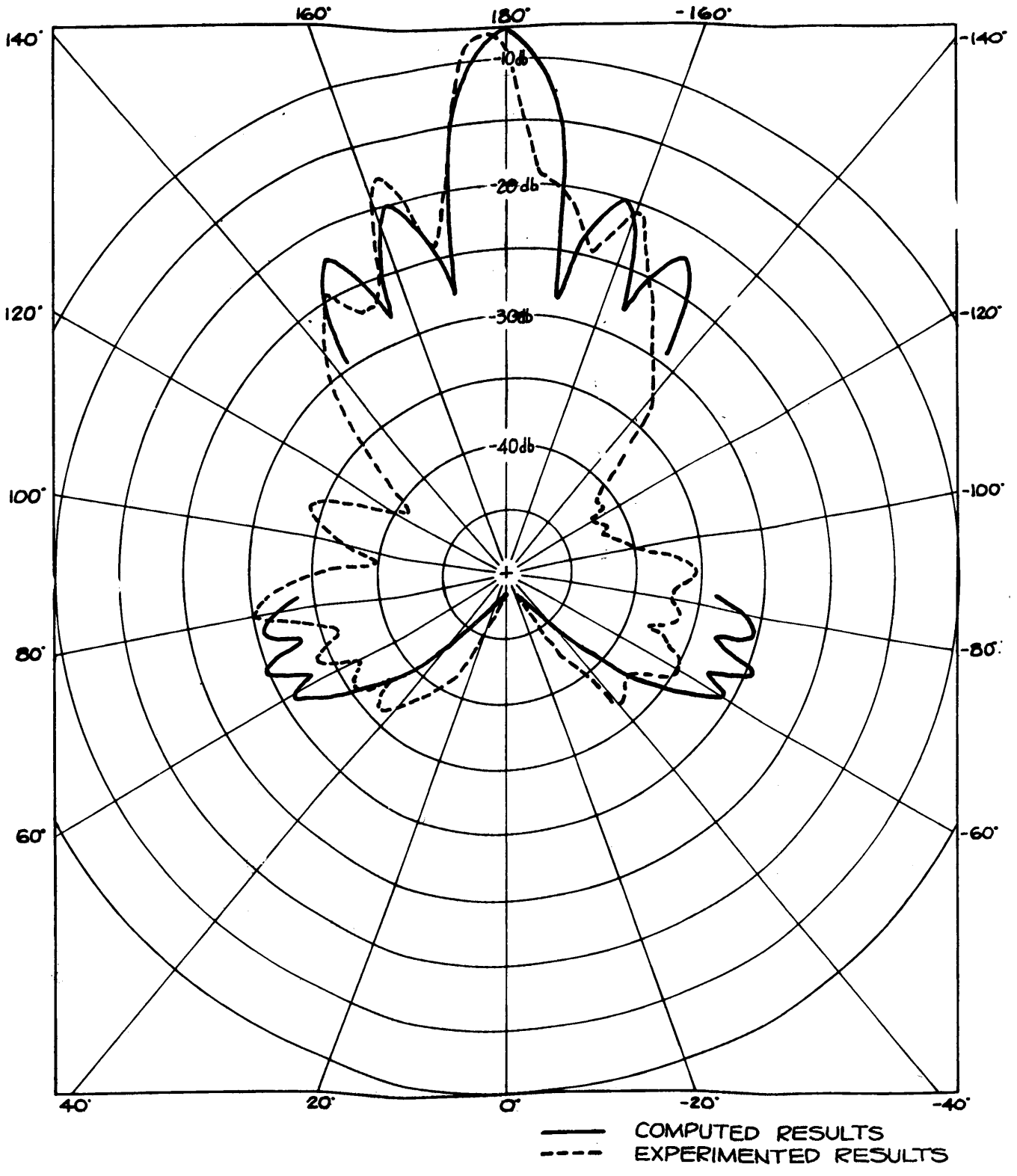


Fig 4-9-DIRECTIVITY PATTERN FOR BAFFLE, INCLUDING TRANSMISSION THROUGH AND DIFFRACTION AROUND BAFFLE

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-33 | 5-26-65 BMB-ES

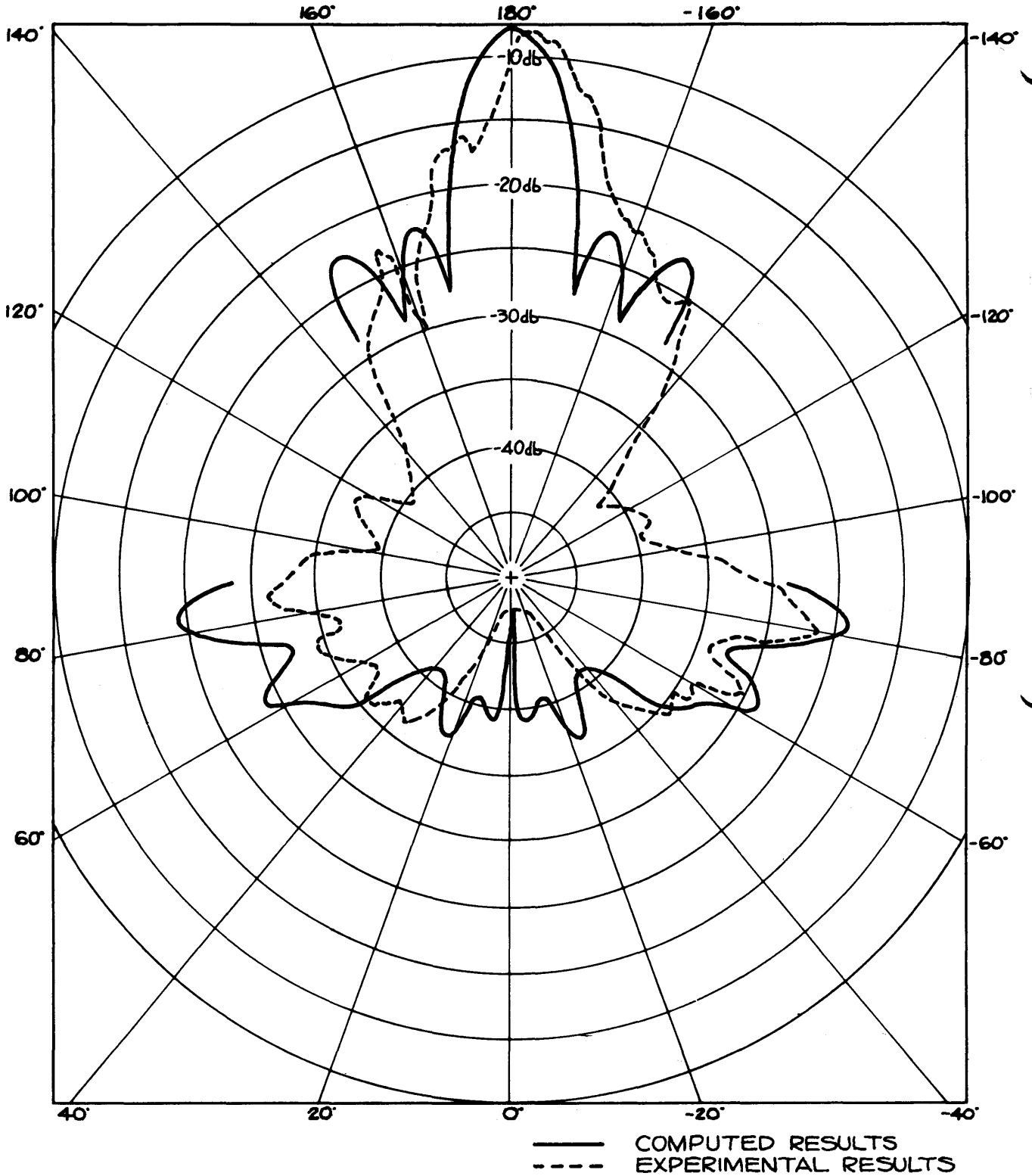
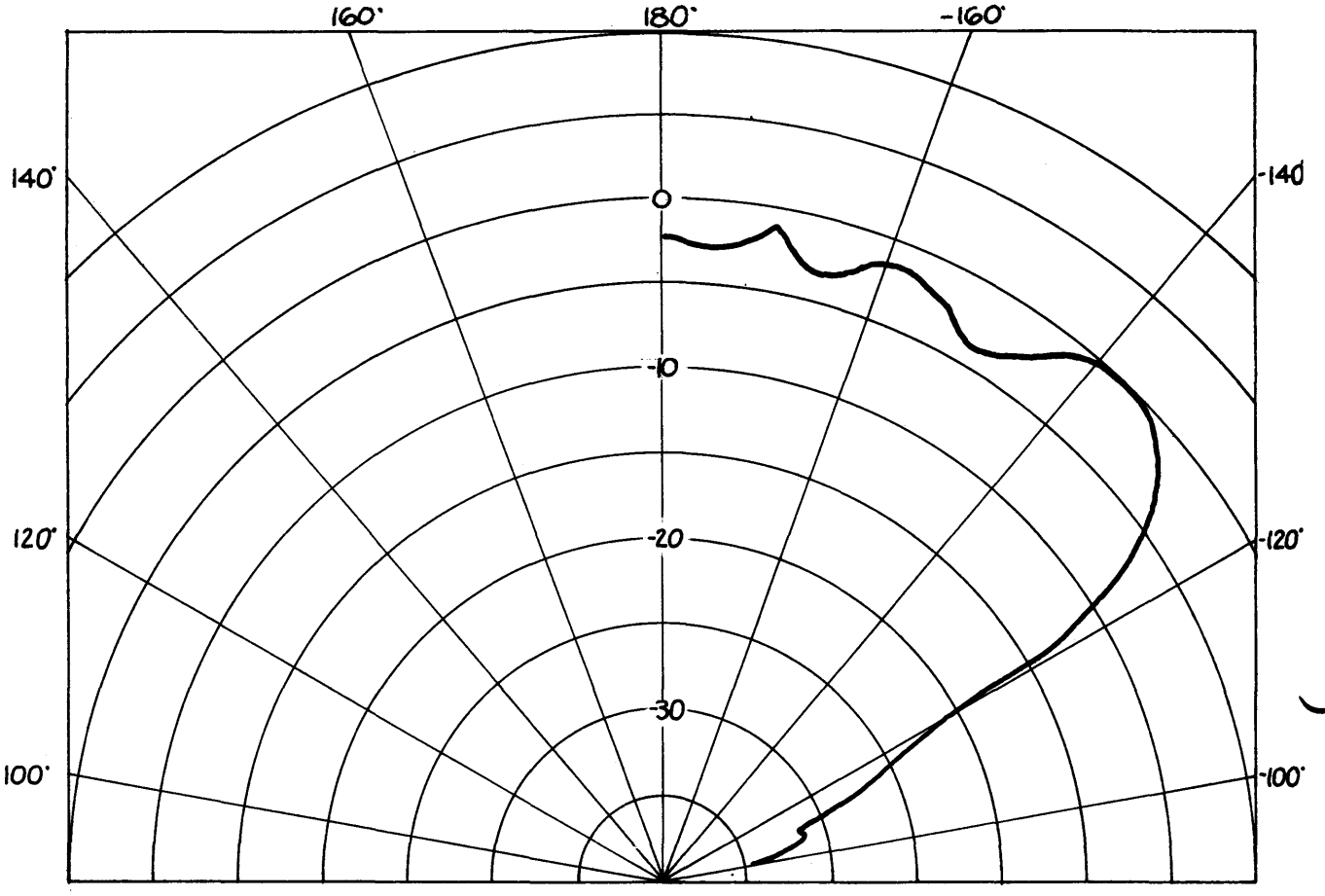


FIG 4-10-DIRECTIVITY PATTERN FOR BAFFLE AND DOME SUPPORTING RODS

TRACOR, INC. | DWG. 771-34
AUSTIN, TEXAS 5-28-65 BMB-ES

to a bare transducer. With a dome, the impedance that the transducer elements "see" is altered. This alteration in radiation loading on the array can change the amount of power which is transferred to the water by the transducer. This loss in signal strength is in addition to the attenuation of the signal due to transmission through the dome. The change in element loading also may affect the element head velocity. Some near-field and far-field sound pressure distributions with and without domes are shown in Figs. 4-11 through 4-14.

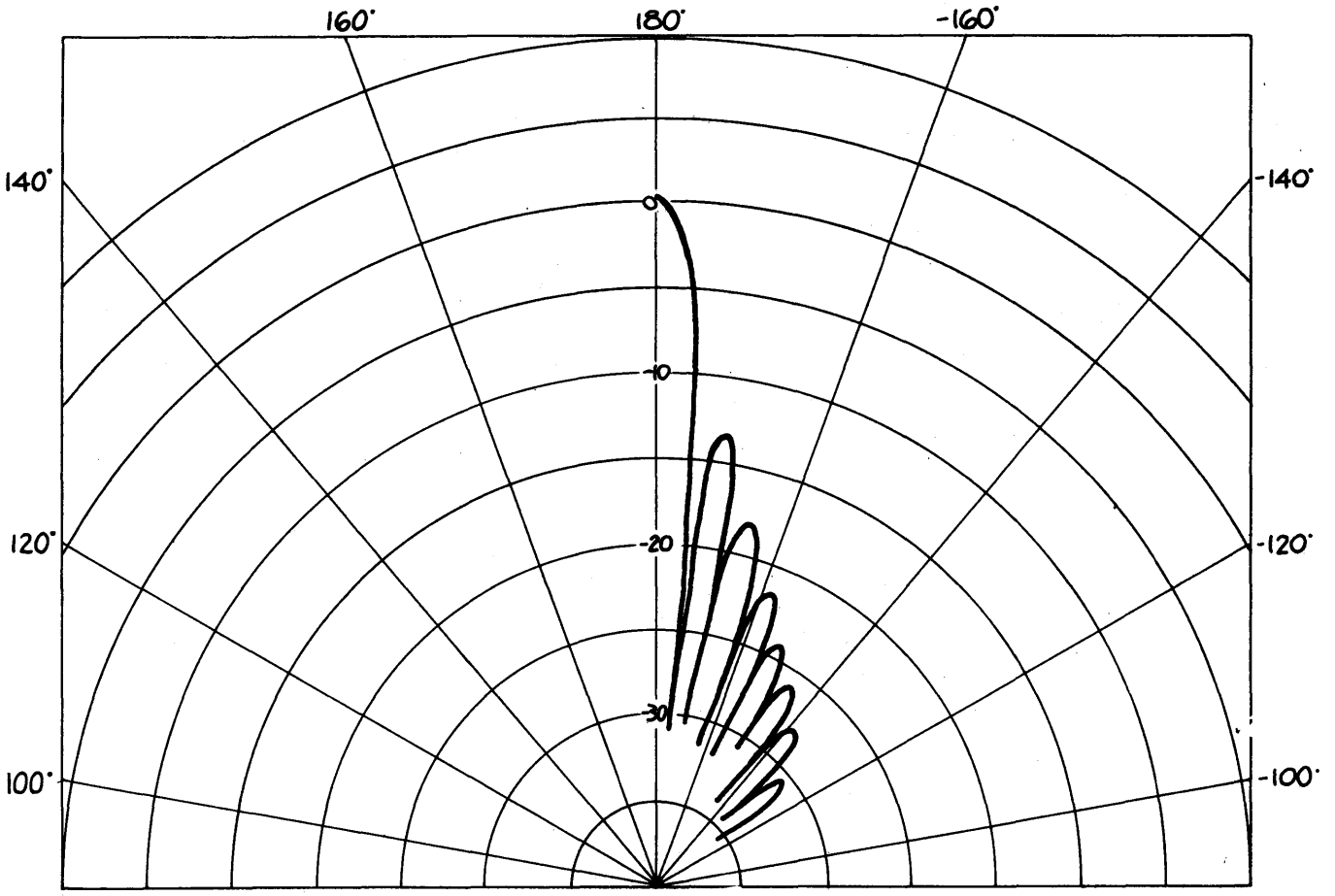


$\frac{\text{CYLINDER RADIUS}}{\text{WAVELENGTH}} = 5.6$

$\frac{\text{RADIUS TO FIELD POINT}}{\text{WAVELENGTH}} = 5.65$

Fig. 4-11-NEARFIELD
NORMALIZED PRESSURE AMPLITUDE (DB):
SECTOR RADIATING IN A RIGID CYLINDER

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-35
5-31-65 BMB-ES

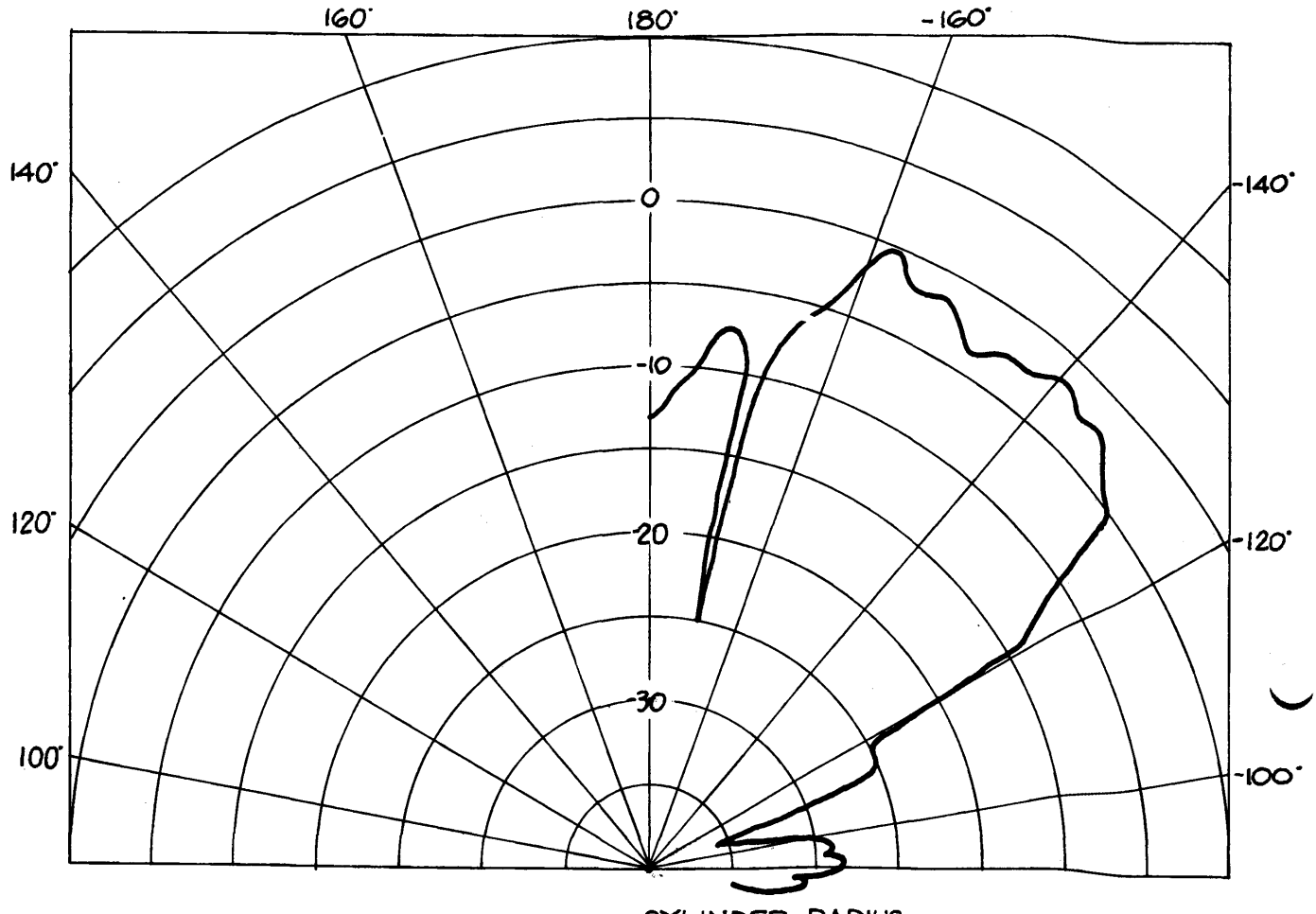


$$\frac{\text{CYLINDER RADIUS}}{\text{WAVELENGTH}} = 5.6$$

$$\frac{\text{RADIUS TO FIELD POINT}}{\text{WAVELENGTH}} = \infty$$

Fig 4-12-FARFIELD
NORMALIZED PRESSURE AMPLITUDE (DB):
SECTOR RADIATING IN A RIGID CYLINDER

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-36 | 5-31-65 BMB-ES



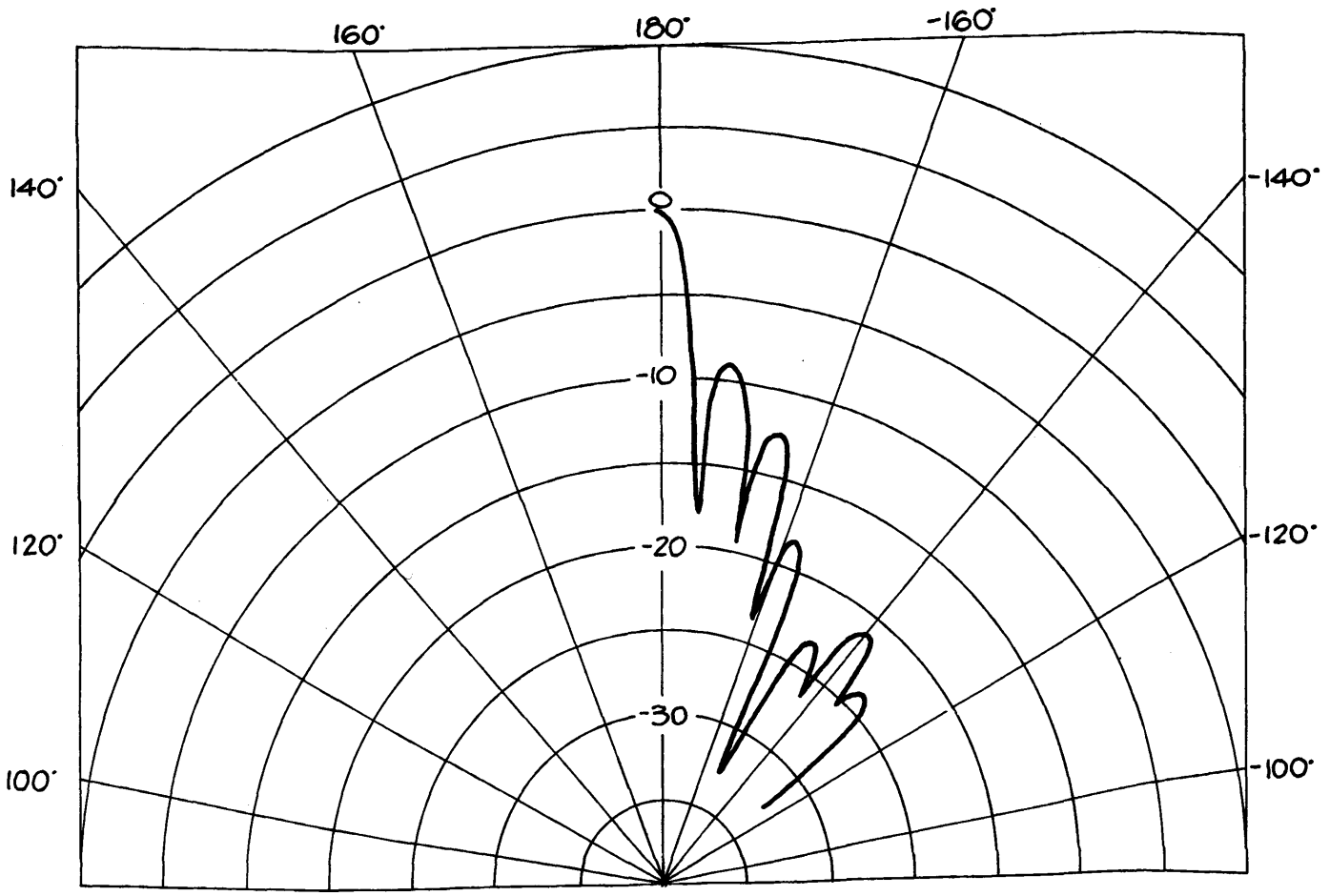
$\frac{\text{CYLINDER RADIUS}}{\text{WAVELENGTH}} = 5.6$

$\frac{\text{SHELL RADIUS}}{\text{WAVELENGTH}} = 7.0$

$\frac{\text{RADIUS TO FIELD POINT}}{\text{WAVELENGTH}} = 5.65$

Fig. 4-13 - NEARFIELD NORMALIZED PRESSURE AMPLITUDE (DB): CYLINDRICAL SECTOR RADIATING INSIDE A CONCENTRIC SHELL

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-37
5-51-65 BMB-ES



$\frac{\text{CYLINDER RADIUS}}{\text{WAVELENGTH}} = 5.6$

$\frac{\text{SHELL RADIUS}}{\text{WAVELENGTH}} = 7.0$

$\frac{\text{RADIUS TO FIELD POINT}}{\text{WAVELENGTH}} = \infty$

Fig. 4-14 – FARFIELD
NORMALIZED PRESSURE AMPLITUDE (DB): CYLINDRICAL SECTOR RADIATING INSIDE A CONCENTRIC SHELL

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-38 | 5.31.65 BMB-ES

5. MILITARY OCEANOGRAPHY

5.1 INTRODUCTION

Oceanography is the science of the ocean. It is subdivided into two parts: (1) Biological, and (2) Physical Oceanography. Physical oceanography is also considered to be a branch of geophysical science.

The importance of the oceans can be noted from the earliest records concerning the best fishing areas and the areas considered to be the best routes to sail. Interest in describing the ocean's gross features, such as waves and tides, flourished with the early classical physicists and mathematicians such as Newton, LaPlace, and Helmholtz and later with Raleigh and Lamb. But oceanography's greatest growth has been during the past 50 years and more precisely since 1941. Two main reasons for the tremendous growth in the last 20 years have been the improvement in instruments and ships and the increase in financial resources directed toward oceanography. As is well known, modern warfare demands as exact a description and as much knowledge as possible about the regions where a conflict is expected. The ocean is by far the most complex battle ground in man's experience. For this reason, so that the many questions influencing military decisions might be answered with the best possible answers, a mighty push has been undertaken to study the ocean environment as well as the ocean's influence on the atmosphere and land masses. The results of oceanographic research that were once considered as "ivory towerish" have been found to apply directly to military problems.

In this section, only a small portion of oceanography can be considered and this portion will be slanted toward the ASW problem in general and underwater sound transmission in particular.

5.2 REFERENCES

Numerous excellent text books and general interest articles are available which treat the different parts of oceanography. The classic text in the field is "The Oceans," by Sverdrup, Fleming and Johnson, published in 1942. Although old in the sense that much has been done since its publication, it still contains the fundamental facts.

The following is a list of texts dealing with the general topics in oceanography.

1. National Defense Research Committee, Military Oceanography, NDRC Div. 6, Vol. 6A.
2. M. N. Hill, The Sea, Vol. I, Phys. Oceanography, Interscience, (1962).
3. A. DeFant, Physical Oceanography, Vols. I & II, Pergammon (1961).
4. C. B. Officer, Underwater Sound, McGraw (1958).
5. R. Weigel, Oceanographic Engineering, Prentice Hall (1964).
6. Hydrographic Office Publication 603, Pierson et al.

5.3 GENERAL CONCEPTS ABOUT THE OCEAN

The oceans cover over 70% of the earth's surface. The depths range downward to about 7 miles. The seas can generate waves with height in excess of 125'. Periods of the waves range from about 1 c/s to 2 cycles/day. The sea contains 330×10^6 cubic miles of water. The volume of all the land above sea level is only about one-eighteenth of the water volume. If all land were smoothed out into one smooth ball, the oceans would cover the earth to a depth of 12,000 feet.

The oceans are divided into four parts: The Pacific, which is larger than the other three combined, the Atlantic, the Indian and the Arctic. Although considered as four separate oceans, they are connected by the vast currents that flow throughout the seas. Though the ocean sometimes appears calm, it is in reality in constant motion due to these large water currents. Several currents, located in the same area, are known to flow in different directions to each other - one at a shallow depth and others at deeper locations.

The oceans and the atmosphere cannot be divorced from each other. The energy exchange between these two is enormous and each influences the other in their general behavior patterns. Atmospheric moisture, and consequently rainfall, depend directly on the ocean. The ocean is a much better heat moderator than the land or the atmosphere so that the seas exert strong control over our weather patterns.

All this activity produces mixing between the different waters but both small scale and large scale variations still exist; the oceans are not homogeneous. The oceans are therefore considered to be nonstationary in both time and space, and this property influences the total ASW program. A few of the more important mechanisms of the ocean and how they influence the ASW problem will be described.

5.4 PHYSICAL OCEANOGRAPHY

5.4.1 The Sea Surface

The surface is described as an acoustic discontinuity (large ratio of acoustic impedance between air and water) which is rough and unstationary. The magnitude of surface roughness effects depend upon the frequency of the sound, the acoustic grazing angle, and the recent history of the surface. The effect of waves on platform (ship) stability is obvious.

There are several systems which make up the sea surface wave pattern.

Tides - the very long period wave motion associated with gravitational forces

Swell - long period waves with periods between 10 and 25 seconds

Sea - waves with periods between 4 seconds and 10 seconds which are locally generated by winds

Capillary - locally generated by wind, propagated by surface tension effects, periods smaller than about 2 seconds.

Recent evidence indicates that ripple state, those waves with periods shorter than about 5 seconds, are of greater importance to the acoustical properties of the surface than longer period waves. This evidence is derived from the Chapman Harris[1] surface scattering curves which correlate wind speed with scattering strength. This relationship implies that locally generated waves are of greatest importance and that longer period waves seem to be of much less importance. Specular reflection of acoustic energy from the sea surface usually has a 4 dB loss, while scattering (which produces reverberation) follows Lambert's Law with a value of -30 dB -10 log $\sin^2\theta$ at a wind speed of about 12 knots. In this formula θ is the grazing angle of the acoustic wave.

Since surface wave motion is a conservative system, the motion of the water particles extends to considerable depth below the surface. In the deep ocean, the water particle motion, which is nearly circular, tends to zero as distance from

---

[1] R. P. Chapman and J. H. Harris, "Surface Backscattering Strengths Measured with Explosive Sources," J. Acoust. Soc. Am., 34, 1592-1597 (Oct., 1962).

the surface increases beyond about one wavelength. These orbital motions act as a strong mixing mechanism and tend to produce temperature mixing so that isothermal surface channel layers can exist to substantial depths.

Ocean waves are classified as either deep-water, surface waves (open ocean) or shallow-water, long waves (near shore). The velocities of the two types are different, depending on the depth d and the wavelength λ and are given by

$$V = \sqrt{\frac{g\lambda}{2\pi}} \quad \text{deep water waves } \left(\frac{d}{\lambda} > \frac{1}{2}\right) ,$$

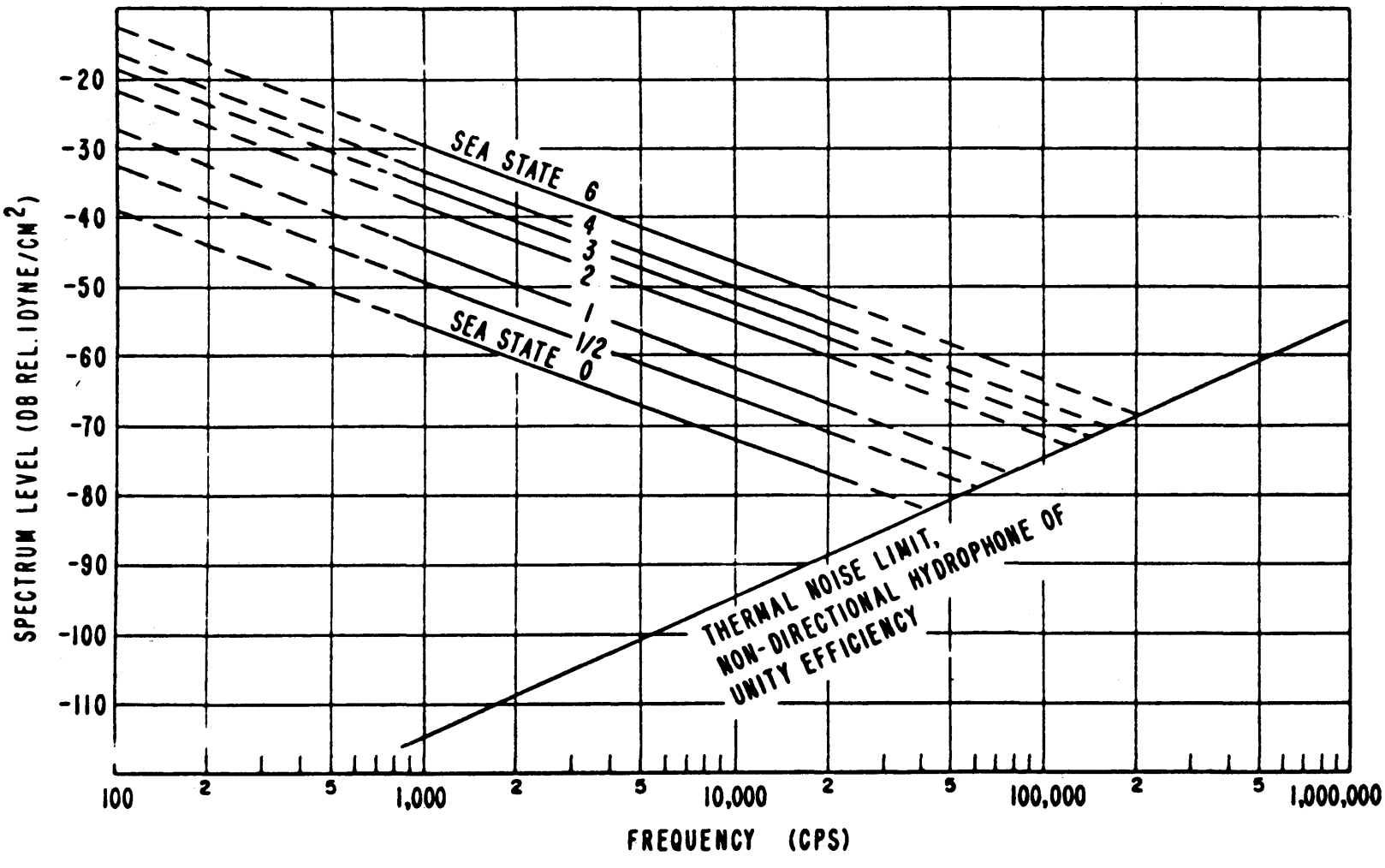
$$v = \sqrt{gd} \quad \text{shallow water waves } \left(\frac{d}{\lambda} < \frac{1}{2}\right) ,$$

g is the acceleration due to gravity and V = wave velocity. An example for two waves considered as typical in the deep ocean $\left(\frac{d}{\lambda} > \frac{1}{2}\right)$ are

Period=T	Velocity=V	Length=λ	Height (typical)=h
6s	10m/s	56m	2m
12s	19m/s	225m	10m

These waves are what would be noticeable on a typical, storm free, day at sea.

A phenomenon associated with wave activity is the production of acoustic noise. Figure 5-1 is a plot of Knudsen Curves which relate deep water ambient noise levels to sea state. This noise is the limiting noise term for sonar systems which are not self-noise limited. The noise is probably due to both wave slap or break as well as to air bubble oscillators trapped near the surface.



DEEP-WATER AMBIENT-NOISE LEVELS (KNUDSEN CURVES) FOR SEA STATES 0 TO 6

FIG. 5-1

TRACOR, INC. AUSTIN, TEXAS DWG. A769-61 11/23/64 JC/MAP

The importance of the sea surface is therefore evident in several different mechanisms influencing underwater sound. The surface influences the scattering mechanisms, the temperature and therefore velocity structure, the ambient acoustic background, and the requirements on sonar design as concerns the platform stability (sonar beam stabilization).

Wind produces waves but the exact energy transfer mechanism is only partly understood. Fetch[2], duration, and speed influence the growth of ocean waves. As a wave enters shallow water, its characteristics change and the orbital motion, in modified form, extends to the bottom. With proper conditions, a near isothermal layer can be developed which extends almost completely to the bottom.

The great ocean currents also have major control of the acoustic conditions. Where the large, well defined currents flow, there are regions of sharp temperature and salinity changes which drastically influence underwater sound propagation. These great streams, such as the Gulf Stream, shed great vortices of water with different acoustic properties which produce regions of unpredictable sonar performance. These currents are related to the average wind patterns over the world as well as to more complex Coriolis effects. Figure 5-2 is a chart of the worldwide ocean flow patterns during the Northern Hemisphere winter season. The mixing of all the ocean waters is clearly evident from this chart. Note that there are numerous instances where large currents flow in opposite directions and are in close proximity to each other.

5.4.2 The Depths

The oceans not only have a large surface area, they are extremely deep. The average ocean depth is taken as 3880 meters (4265 yards). The relationship between the land and ocean heights

[2]The distance over which wind direction is constant.

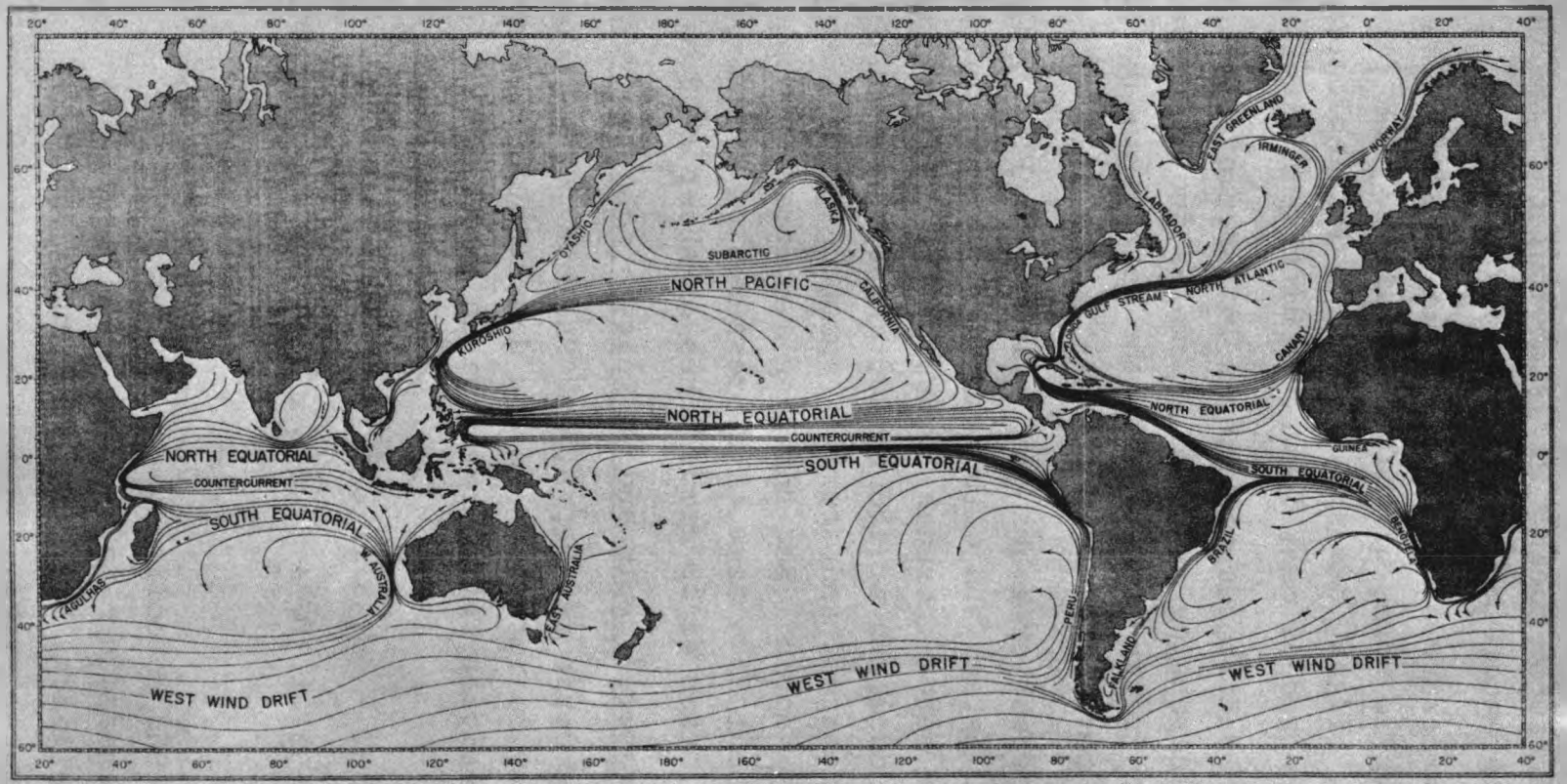


FIG. 5-2 —OCEANIC SEA SURFACE CURRENTS

SOURCE: MILITARY OCEANOGRAPHY
NDRC DIV. 6,
SUMMARY TECHNICAL REPORT
VOL. 6A

TRACOR, INC. | DWG. 6367
AUSTIN, TEXAS 11/29/65 AG/SL

and depths is shown in Fig. 5-3. Figure 5-4 shows a cross section of the Atlantic Ocean. It is obvious that the ocean depths are much greater than the land elevations.

The structure of the ocean floor is complex and strange to man's thinking in terms of what is familiar. The floor consists of plains, ridges, deeps, and slopes. The geological structure is composed of many layers of different materials.

Depth contributes to changes in underwater sound propagation since increasing pressure increases the sound velocity. Although large currents are present, the deep ocean approaches a near constant temperature.

The deepest trenches (as indicated in Fig. 5-4) are in excess of 10,800 meters and with a nominal increase in pressure of 1 decibar per meter increase in depth, the maximum pressure is about 10,000 decibars or about 1300 atmospheres. Only the most carefully designed and sturdily constructed apparatus can operate at these depths. But for even a moderate depth, say 700 meters or about 2100 feet, the pressures on a submarine hull would be of the order of 65 atmospheres or about 1000 psi.

5.4.3 Temperature, Depth, and Salinity - Acoustic Velocity

The Temperature, Depth, and Salinity (TDS) dependence of sound velocity is important in the description of sonar behavior. The velocity of sound is given by

$$c = 1448.6 + 4.63T - (5.38 \times 10^{-2})T^2 + (3.54 \times 10^{-4})T^3 + (S-35)\,1.307 + (1.815 \times 10^{-2})D - (S-35)(1.07 \times 10^{-2})T,$$

where

c = Sound velocity, yards/second
T = Temperature in ^{o}F
S = Salinity, Parts per thousand, $^{o}/oo$
D = Depth, yards.

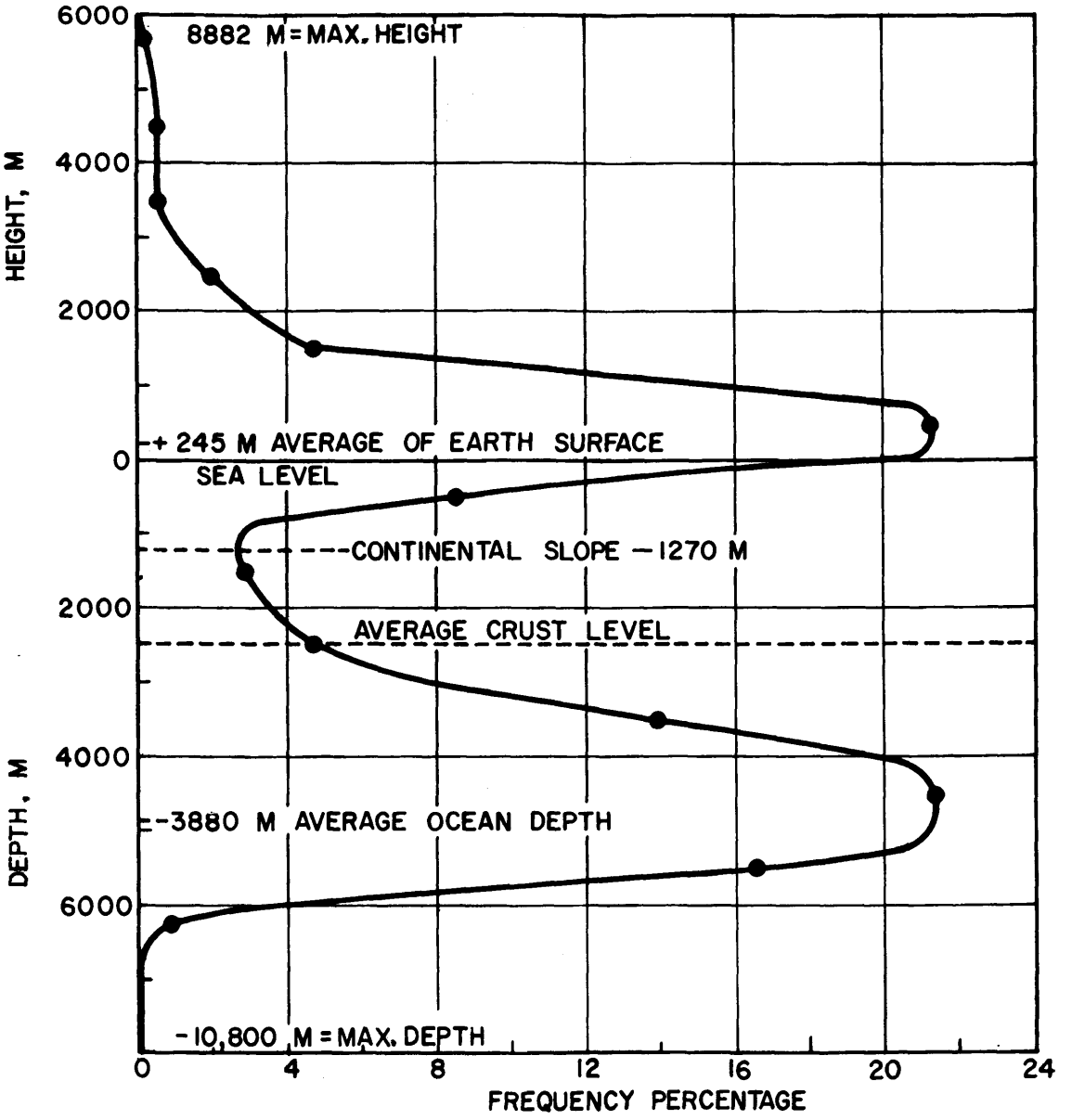
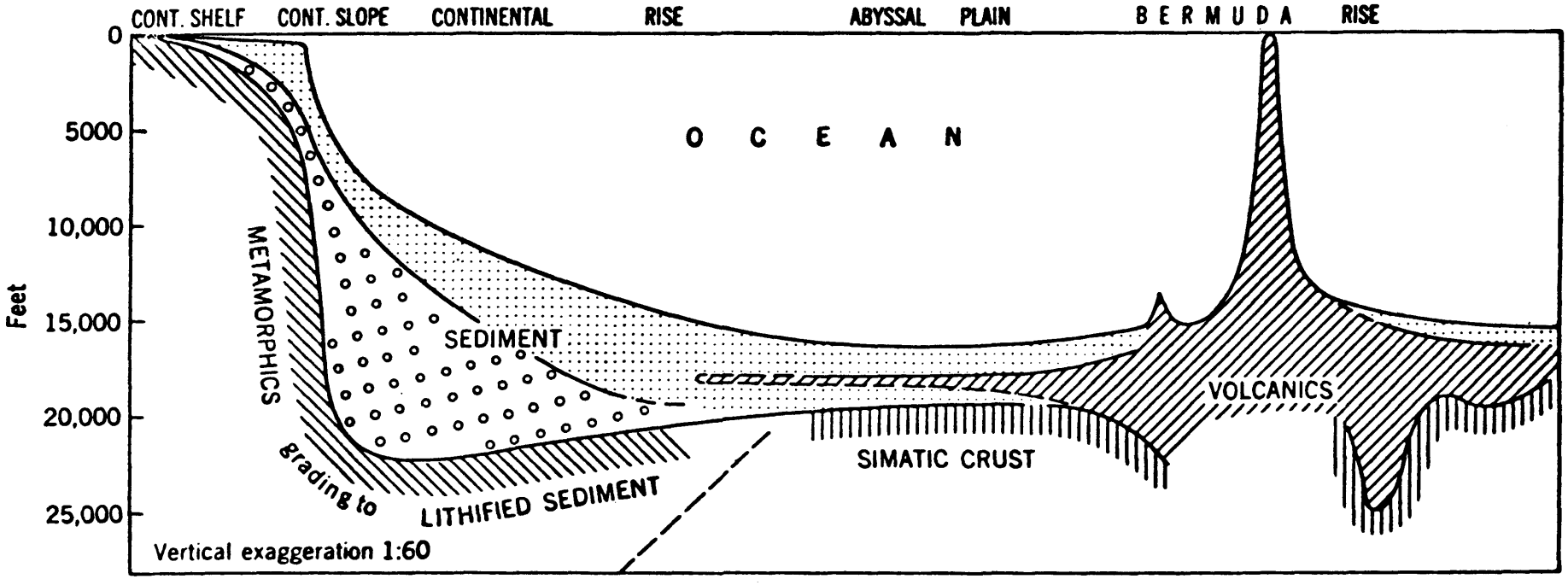


FIG. 5-3 - DISTRIBUTION OF HEIGHT AND DEPTH INTERVALS OVER THE EARTH'S SURFACE.

TRACOR, INC. | DWG. 769-67
AUSTIN, TEXAS | 5-25-65 | JC-ES



(From C. B. Officer, Deep Sea Research, 2, 1955.)

FIG.5-4 -TYPICAL PROFILE OF THE OCEAN FLOOR

TRACOR, INC. AUSTIN, TEXAS DWG. A769-65 11/24/64 JC/MAP

A somewhat less cumbersome relationship, although a rougher approximation, is given by

$$c = 1200\ (1 + 0.2 \log T) + (S-35)(2.3 - 0.6 \log T) + 0.018D,$$

where the terms are defined as before.

The Wilson Tables[3] provide highly precise values of sound velocity as a function of temperature, salinity, and pressure.

Temperatures in the oceans range from -2°C to +30°C (28°F to 90°F). The temperature structure is greatly variable, both in a given area, from area to area, and as a function of time. A typical temperature profile, along with an indication of how it varies with time, is shown in Fig. 5-5. Figures 5-6 and 5-7 are measured curves in the areas indicated and show how the temperature and velocity profiles change between measurements. The great bend in the velocity profile curves near the 1200 yard depth is the SOFAR channel where sound energy is trapped and propagates over great distances with small attenuation.

Pressure increases with depth at a rate described previously. Sound velocity also increases with depth at a rate of 0.018 yards/second per yard increase in depth. This term becomes important because of the great depths possible.

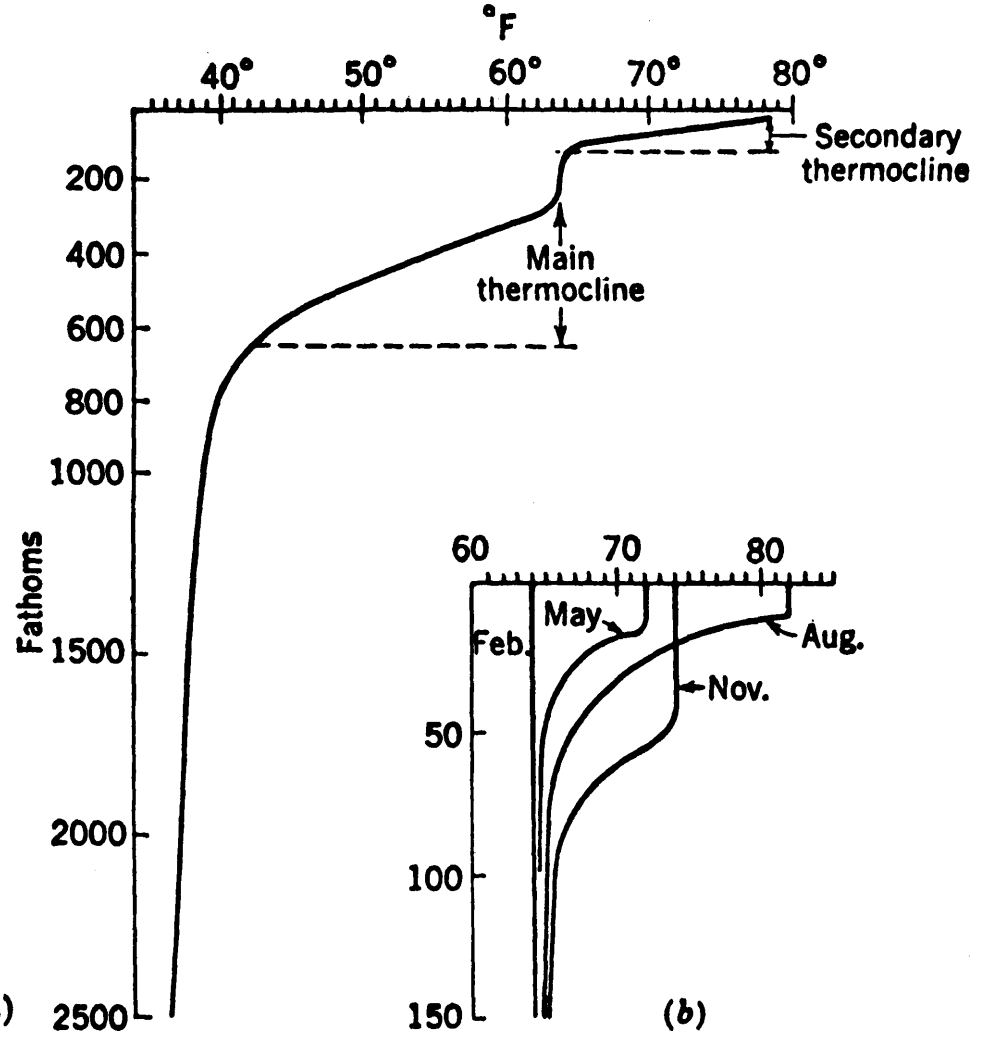
A summary of the velocity equation can now be made in light of the values of temperature, salinity and depth.

a. At a nominal temperature of 60°F and salinity of 35°/oo, the temperature coefficient of velocity change is about 2.17 yards/second per degree change in temperature (near the surface).

b. At a nominal temperature of 60°F and salinity of 35°/oo, the salinity coefficient of velocity change is about 1.42 yards/second per part per 1000 change in salinity (near the surface).

---

[3]U. S. Naval Oceanographic Office, "Tables of Sound Speed in Sea Water," Special Publication SP-58, Washington, D. C. (Aug., 1962).



(From Woods Hole Oceanographic Institution, Natl. Defense Research Council, 1941.)

FIG. 5-5- TYPICAL TEMPERATURE PROFILE

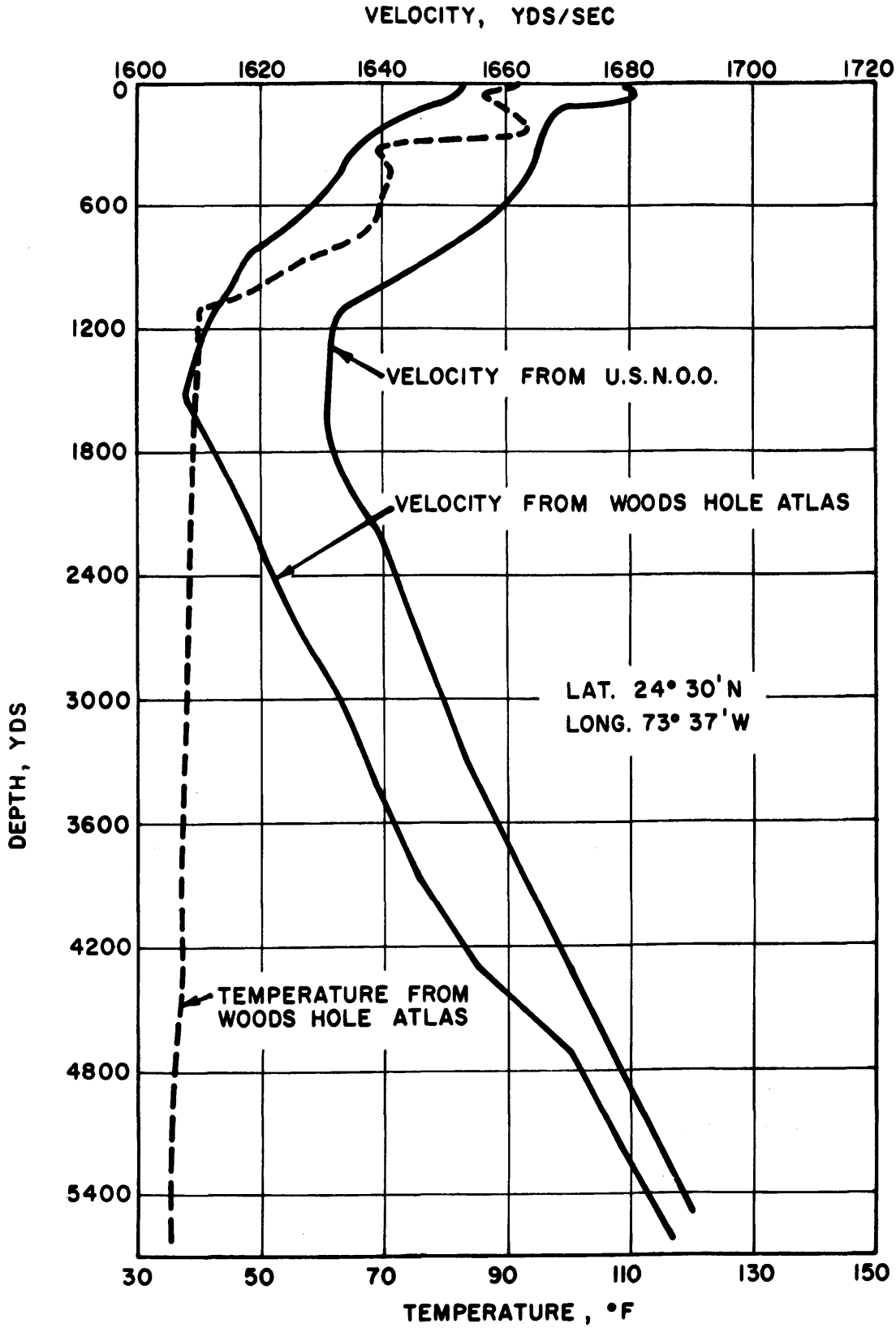


FIG. 5-6 — DEEP OCEAN TEMPERATURE AND SOUND VELOCITY PROFILES

TRACOR, INC. AUSTIN, TEXAS
DWG. A 769-1
10-12-64 JLC-MM

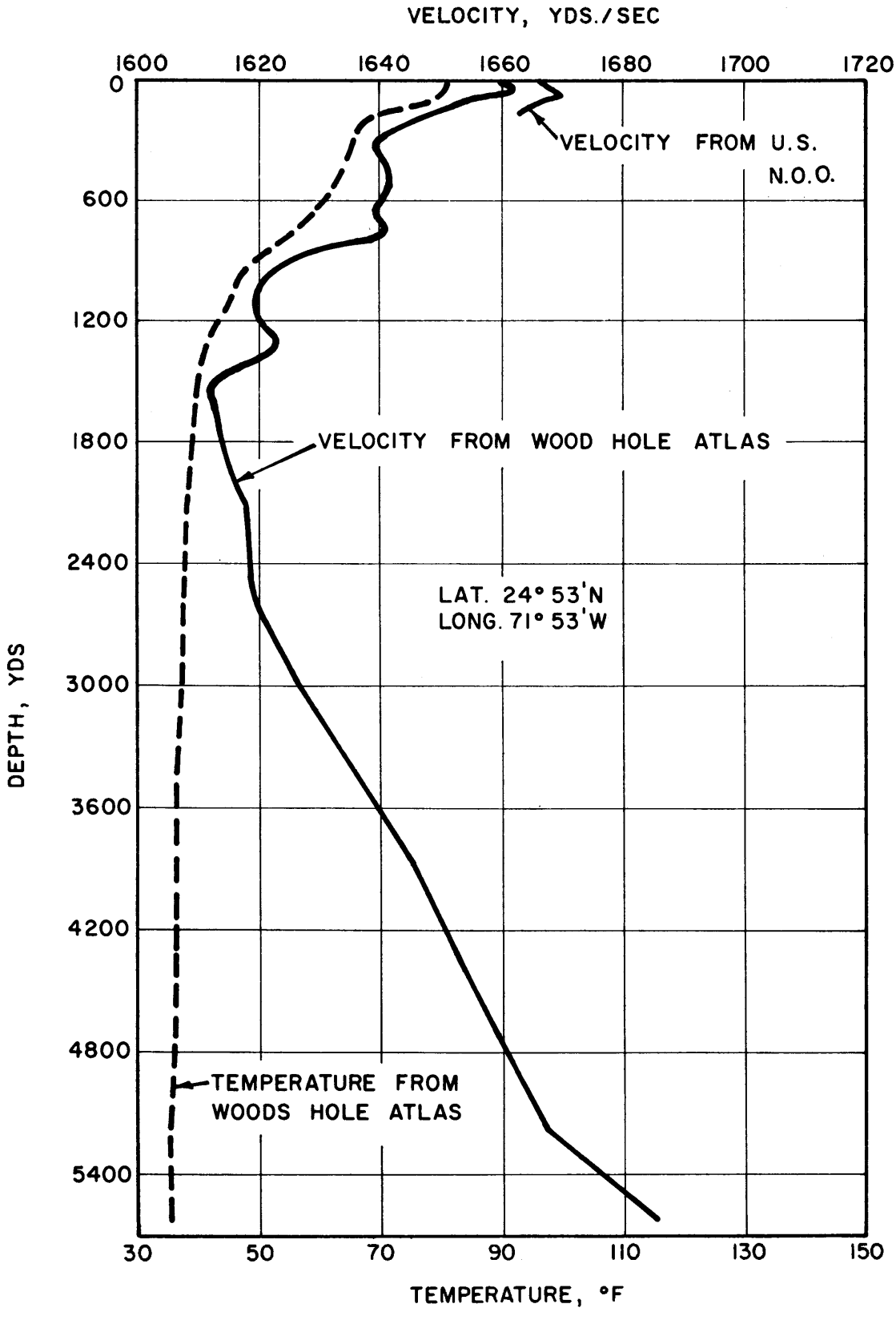


FIG. 5-7 –DEEP OCEAN TEMPERATURES AND SOUND VELOCITY

TRACOR, INC. DWG A 769-2
AUSTIN, TEXAS 10-12-64 JLC-MM

c. The change in sound velocity with depth is 0.018 yards/second/yard increase in depth.

Therefore, temperature dominates in the control of sound velocity. But in regions where the temperature becomes constant, depth tends to control the velocity. Because of the great depths and the influence of depth on the velocity profile, sound is refracted upward and the convergence zone propagation mode is possible.

5.4.4 Internal Waves

The internal wave is a wave propagating along a boundary between regions in the water of different density, similar to surface waves. Internal waves are normally associated with temperature oscillations in the main thermocline region (see Fig. 5-5). But the temperature oscillations are merely an indication of the phenomenon and a convenient method of measuring the effect.

Internal waves have periods of from 5 to 10 minutes with components out to 30 minutes and beyond. Their velocity is low compared to surface waves and is about 0.5 to 1.0 ft/second. Wave lengths therefore amount to several hundred feet. There are periods of 12 and 24 hours associated with tides. The velocities of internal waves in deep water and shallow water are given by equations similar to these for surface waves. The velocity expressions are

$$V = \sqrt{\frac{g\lambda}{2\pi}\frac{\rho - \rho'}{\rho + \rho'}}$$ Deep water waves $\left(\frac{d}{\lambda} > \frac{1}{2}\right)$

$$V = \sqrt{gd\,\frac{\rho - \rho'}{\rho}}$$ Shallow water waves $\left(\frac{d}{\lambda} < \frac{1}{2}\right)$

where ρ' is the density of the upper fluid layer, ρ is the density of the lower fluid layer, λ is the wavelength and V is the velocity. The constant g is the acceleration due to gravity and d is the water depth.

Internal waves produce a reflecting and refracting layer that is unstationary. Acoustic properties through the layer change because of the change in grazing angle a ray experiences from time to time. These waves also seem to contribute to amplitude fluctuation of energy propagated in their neighborhood. Since particle motion is associated with these waves, it is also suspected that they contribute to micro-structure phenomena which lead to scattering and reverberation. It also seems that the so-called "deep scattering layer" is associated with this thermocline region.

5.5 BIOLOGICAL OCEANOGRAPHY

Biological mechanisms influence ASW programs in two main ways: (1) scattering of acoustic energy; (2) plant and animal growth on ship hulls and transducer components.

Life in the ocean ranges from microscopic sizes up to the whales 100 feet in length. Animal life is found at all depths from the surface into the bottom at the deepest trenches. Animals may be found alone or in groups or schools. The basic life in the ocean is plankton - from which all others ultimately feed - small fish and animals consume plankton and the larger fish consume the smaller.

But the real interest in this discussion is in the biological effects on acoustic energy. The deep scattering layer is an example of a scattering mechanism that is not totally understood. It seems to rise and fall in depth, depending on the time of day. The layer rises at night and descends in the day time. The exact mechanism has not been identified since nets pulled through the layer do not "catch" anything which

could be identified as a scatterer. Photographs indicate fish near the layer but not in sufficient quantity to produce this effect.

Other marine life forms are known sources of noise and interference. Snapping shrimp, groupers and whales have been identified as very noisy sources, and the target strength of a whale or a large fish is sufficient to produce echoes. The abundance of marine life located at almost any depth, but primarily in the upper 1000 ft., contribute to volume reverberation as individual scattering mechanisms.

Marine life has been and is presently contributing to geological formations and changes in the bottom topology. It also contributes to food, agriculture, and future oil and coal resources.

5.6 CHEMICAL OCEANOGRAPHY

Through the study of ocean chemistry, it has been determined that the oceans are about 3 billion years old. Their chemical evolution is a continuing process.

The oceans contain enormous quantities of dissolved and precipitated minerals. But of primary interest to ASW is the salt concentration and its effects on sound propagation.

In terms of the basic elements, the impurities of the ocean are composed of chlorine - 55%, sodium - 30%, sulphates - 7.7%, magnesium - 3.7%, calcium - 1.2%, and potassium - 1.1%. Even normally rare manganese and gold are present in significant quantities. Exact percentages vary from region to region.

The high concentration of these active ions in sea water makes it one of the most corrosive solutions in nature. This corrosive nature presents one of the major obstacles to designing and maintaining sophisticated equipment needed in modern ASW operations.

5.7 GEOLOGICAL OCEANOGRAPHY

ASW interest in geological oceanography is primarily concerned with the acoustic reflection and transmission characteristics of the ocean floor. Although sediment transport and structure must be considered in designing ASW systems which will be installed or anchored on the bottom, these effects are usually treated as ocean engineering problems.

Acoustic energy propagation involves the bottom in both shallow (depth < 100 fathoms) and deep water. In shallow water, a bounded channel is formed between the sea surface and the bottom. Acoustic properties of the bottom are therefore important in predicting sonar performance.

In considering the effects of the bottom on scattering (both specular and non-specular), bottom material, sub-structure and topography are the primary controlling features. Material ranges from mud to sand to rock and the acoustic properties vary accordingly. Topography ranges from smooth (abyssal plains) to very rough (sea mounts and canyons). Sub-structure varies from many closely spaced reflecting layers to the case of only one continuous layer (to the penetration depth at the frequency used in measurement). A resume of bottom effects can be found in a recent unclassified publication titled "Underwater Reverberation as a Factor in ASW Acoustics," by Peter A. Barakos, USNUSL Report #620 dated 11 September 1964.

To summarize pertinent geological effects on acoustic properties:

a. Bottom loss (specular, coherent reflections). In the 3.5 kcs frequency range and at grazing angles greater than 15^{o}, the bottom loss will be approximately 12 dB for the Atlantic and 15 dB for the Pacific. These are mean values; selected areas could exhibit losses somewhat different from these values.

Bottom loss is $B = 10 \log \frac{I_r}{I_i}$ where I_r = reflected intensity and I_i = incident intensity. Bottom loss is a function of frequency and grazing angle.

b. Scattering (reverberation). Scattering coefficients for the bottom have a range of values from -35 dB to -10 dB with a mean value near -28 dB. Scattering seems to follow a Lambert's Law scattering relationship ($\alpha \sin^2 \theta$) in certain special cases. As with bottom loss, the scattering coefficient is considered as $\mu = 10 \log \frac{I_s}{I_i}$ where I_s is the scattered intensity and I_i is the incident intensity, at a given grazing angle.

6. SIGNAL PROCESSING

6.1 BACKGROUND

Before beginning the formal discussion of signal processing it will be instructive to consider a familiar signal processor, the human ear. It turns out that its behavior can be described as equivalent to the behavior of a filter network. At this place in the discussion an experiment will be described; an interpretation of the result will be given in a later section as an example of the application of signal processing principles.

Experiment: The output of a Gaussian noise generator adjusted to an average noise power level N was recorded on magnetic tape in 15-second sections. Steady sinusoidal signals having duration τ and at average power S were embedded in half of the 15-second noise sections, or "echo" cycles, and listeners were asked to listen to all of the "echo" cycles and respond by pushing a button whenever they thought they heard a signal. (The listeners had been trained before hand on easily detectable signals; the signal levels were gradually reduced to the levels employed in the experiment.) A "hit" was scored for a listener when he responded within one second after the point where an actual signal occurred. A "false alarm" was scored for a listener if he responded in any noise section in which a signal had not been embedded. The listeners were not restricted to one response per noise section. Scores were tabulated using the last response in each noise section. Two ratios were computed.

$$D = \frac{\text{Number of "Hits"}}{\text{Number of Noise Sections with Signals}}$$

$$F = \frac{\text{Number of "Empty" Sections in which Responses Were Made}}{\text{Number of "Empty" Noise Sections}}$$

An "empty" noise section is a noise section in which no signal

was embedded. The ratio D is an estimate of the detection probability. The ratio F is an estimate of the probability of making at least one false alarm in a 15-second noise section.

After the listeners had listened to several hundred noise sections on different days and under different instructions concerning how strict or relaxed they should be about responding, it was found that their detection probability D plotted against their probability F of making at least one false alarm per 15-second "echo" cycle grouped around the line marked $\frac{S}{N} = 0$ dB in Fig. 6-1. The ratio of average signal power to the average noise power is usually expressed in decibels (dB). Written in these units, it is

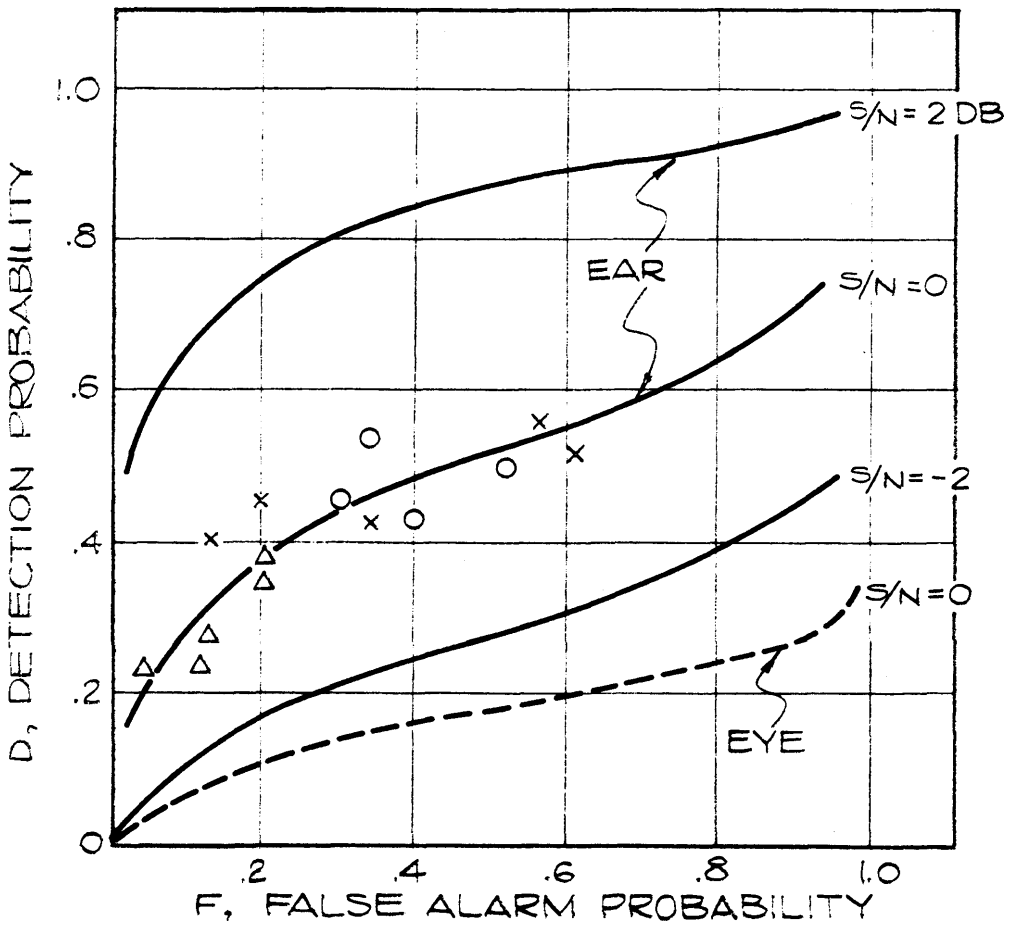
$$\frac{S}{N}(\text{dB}) = 10 \log \left[\frac{\text{Average Signal Power (Watts)}}{\text{Average Noise Power (Watts)}}\right].$$

Since $\log(1) = 0$, $\frac{S}{N} = 0$ dB means that the average signal power is equal to the average noise power.

The solid line marked $\frac{S}{N}=0$ in Fig. 6-1 was computed theoretically. The listener responses, where the average signal power/average noise power was raised to $\frac{S}{N} = +2$ dB, clustered about the upper solid line. They clustered about the lower solid line when the signal-to-noise ratio was lowered to $\frac{S}{N} = -2$ dB.

These plots indicate that the listener's probability of making a detection at a given false alarm probability is increased when the signal-to-noise ratio is increased. The results may also be interpreted by saying that the probability of making a mistake is decreased by raising $\frac{S}{N}$, if operation at a given detection probability is maintained. This experiment indicated that improvement of the signal-to-noise ratio aids the listener in making a correct decision about the presence or absence of the signal.

The signal-to-noise ratio is not the whole story however. Suppose the same noise sections are written out as pen recordings



T = 15 SEC τ = 70 ms

(S/N)$_i$ = 0 DB B_{NOISE} = 210 C/S

△ OBSERVER 2
○ OBSERVER 3
× OBSERVER 4

FIG.6-1- OBSERVER ROC

TRACOR, INC. | DWG. 771-41
AUSTIN, TEXAS | 6-23-65 BMB-ES

and the same listeners were asked to detect the signals by eye in the recorder charts. It will be found that for the signal-to-noise ratio $\frac{S}{N} = 0$, the plot will be lower (see the dotted curve marked $\frac{S}{N} = 0$ in Fig. 6-1). This suggests that the eye is not as good a processor as the ear and emphasizes the fact that a knowledge of the signal-to-noise ratio is not the whole story.

The curves shown in Fig. 6-1 are known as "Receiver Operating Characteristics" (ROC). These curves are much more suitable for comparing performances of signal processing systems than are signal-to-noise ratios. Further discussion of ROCs and their prediction will be given in a later section.

6.1.1 The Sampling Theorem

The first step in setting the stage for the signal processing discussion is to try to establish what are known as sampling theorems. Most time functions which must be described exhibit a variability and complexity which is difficult to describe. Even a simple constant amplitude sinusoidal function of finite duration, if described point by point, is cumbersome. There is obviously a convenient description for this function because five quantities suffice, namely,

A - amplitude
ϵ - epoch
f - frequency
T - duration
t_1 - beginning of pulse

These quantities, when inserted into the expression

$$u(t) = A \sin(2\pi f_o t + \epsilon) \qquad t_1 \leq t \leq t_1 + T \qquad (6\text{-}1)$$

$$= 0 \qquad \text{otherwise}$$

completely specify this particular time function.

The sampling theorem establishes the minimum number of quantities needed to describe an arbitrary time function. Suppose a band limited time function u(t) having duration T is to be represented by a Fourier series. The form of the series is

$$u(t) = \frac{a_o}{2} + \sum_{h=1}^{\infty} [a_h \cos 2\pi h f_1 t + b_h \sin(2\pi h f_1 t)] \quad (6\text{-}2)$$

in which the a_h and b_h are given by

$$a_h = \frac{2}{T} \int_0^T u(t) \cos(2\pi h f_1 t)\, dt$$

$$b_h = \frac{2}{T} \int_0^T u(t) \sin(2\pi h f_1 t)\, dt$$

where $f_1 = \frac{1}{T}$ is known as the fundamental frequency.

Because u(t) is band limited to a bandwidth B, the a_i and b_i corresponding to frequencies hf_1 outside the bandwidth B will be identically zero. This situation is pictured in Fig. 6-2. The lowest frequency present in u(t) is shown as kf_1 and the highest frequency present is shown as $(k + j)f_1$. In the representation of u(t) by Eq. (6-2) those terms prior to h=k and following h = k + j may be deleted from the sum with the result that u(t) is represented by the finite series,

$$u(t) = \sum_{h=k}^{k+j} [a_h \cos(2\pi h f_1 t) + b_h \sin(2\pi h f_1 t)] \quad (6\text{-}3)$$

This series provides an exact representation of u(t) over the interval of length T if u(t) is continuous.

The question may now be asked as to how many quantities need be specified to describe u(t). There is first $f_1 = \frac{1}{T}$, the reciprocal of the duration of the time function. Then there are (j + 1) of the a_i and (j + 1) of the b_i, a total of

$$\eta = 2(j + 1) \tag{6-4}$$

coefficients.

This number is clearly related to the bandwidth of u(t) because the bandwidth is divided into at least j intervals of width f_1, hence the equality,

$$jf_1 = B, \tag{6-5}$$

must hold. Elimination of f_1 from Eq. (6-5) in terms of T and use of the value of j given by Eq. (6-4) gives the result that the number of quantities which must be specified in the description of u(t) satisfies the equality,

$$\eta = 2(BT + 1) \tag{6-6}$$

The number η does not include the quantities kf_1 and T. In other words the number of quantities needed to specify a time function u(t) having bandwidth B and duration T is (2BT + 2).

Equation (6-6) is of great practical importance. It alone determines the number of digital samples which must be made when processing u(t). Since 2BT + 2 constants must be determined, 2BT + 2 independent samples of u(t) must be obtained. Suppose that $u(t_1)$, $u(t_2)$, ---, $u(t_\eta)$ equally spaced values of u(t) are read from the graph of u(t) and η simultaneous equations of

the form,

$$u(t_1) = \sum_{i=k}^{k+j} [a_i \cos 2\pi\, if_1 t_1 + b_i \sin 2\pi\, if_1 t_1]$$

$$u(t_2) = \sum_{i=k}^{k+j} [a_i \cos 2\pi\, if_1 t_2 + b_i \sin 2\pi\, if_1 t_2] \qquad (6\text{-}7)$$

$$\vdots$$

$$u(t_\eta) = \sum_{i=k}^{k+j} [a_i \cos 2\pi\, if_1 t_\eta + b_i \sin 2\pi\, if_1 t\eta]$$

are written. The equally spaced samples of u(t) would completely specify all of the a_i and b_i. The complete original time function could be reconstructed from Eq. (6-3) and the a_i and b_i which were found with Eqs. (6-7). It is implied by this discussion that the equally spaced samples represent a complete set of independent samples of u(t) and that additional samples contain no new information about the sampled function.

On the strength of the preceding statements it is concluded that only a number η of independent samples of u(t) can be taken in a time T and that η is 2(BT + 1). This statement is sometimes made in terms of the number of independent samples which may be made per unit time. If Eq. (6-6) is divided by T, the equality becomes

$$\dot{\eta} = 2B + \frac{2}{T}\,, \qquad (6\text{-}8)$$

in which $\dot{\eta}$ means the number of independent samples which may be made per second. This result is familiar. Just over 2B samples per second may be considered to be independent.

The discussion to this point has been made under the tacit assumption that a narrow-band time function was under discussion. Figure 6-2 shows a function whose bandwidth is smaller than its center frequency. A wideband time function, one for which the bandwidth is comparable to the center frequency, is shown in Fig. 6-3. For this case, as in the narrow-band case,

$$jf_1 = B \ , \tag{6-9}$$

The same arguments which lead to Eq. (6-6) can be made again.

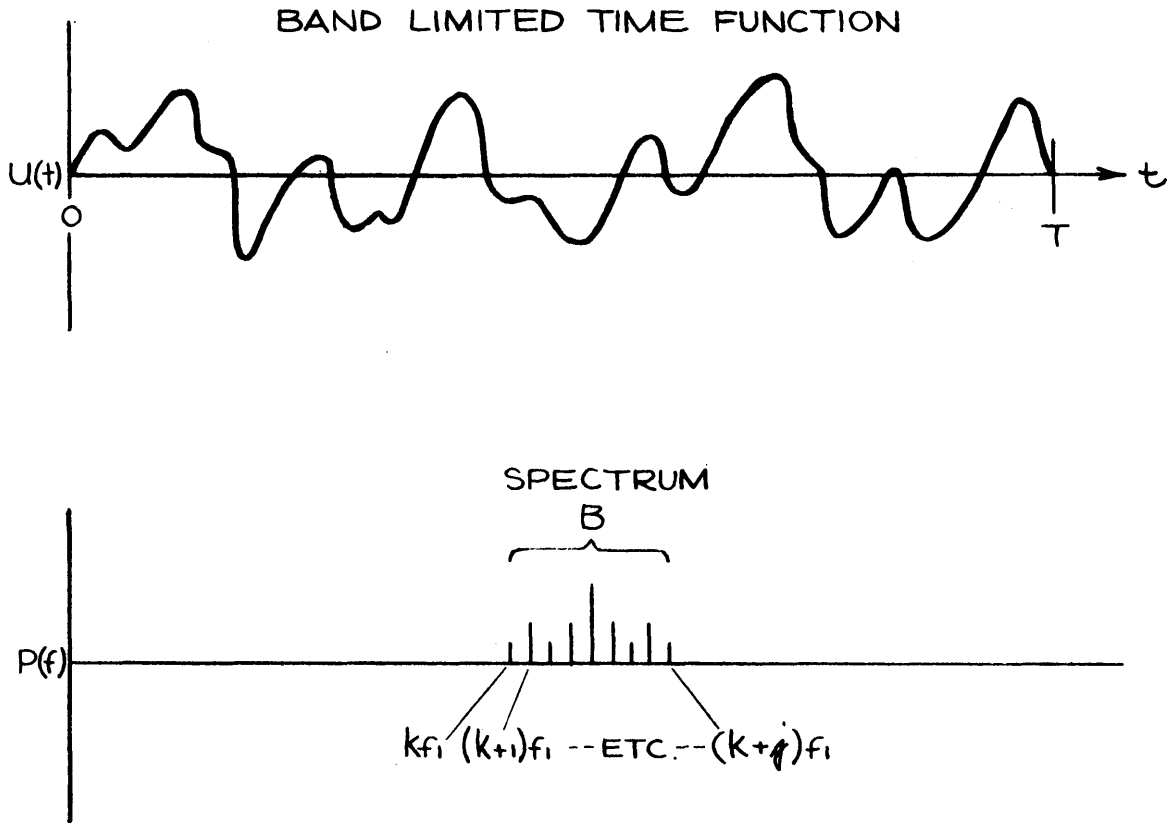
Sampling is discussed further in a number of references. Woodward[1] presents a readable account. An important requirement described there in sampling narrow-band time functions is that the samples needed are the amplitude and simultaneous phase at equally spaced sample points. $2B + \epsilon$ ($\epsilon > 0$) samples are needed each second but only $B + \frac{\epsilon}{2}$ amplitude samples are needed. This means that only $(B + \frac{\epsilon}{2})$ independent power samples can be obtained per second.

Theorem I:

> A band limited wide-band time function is completely specified by $2B + \epsilon$ equally spaced samples per second ($\epsilon > 0$).

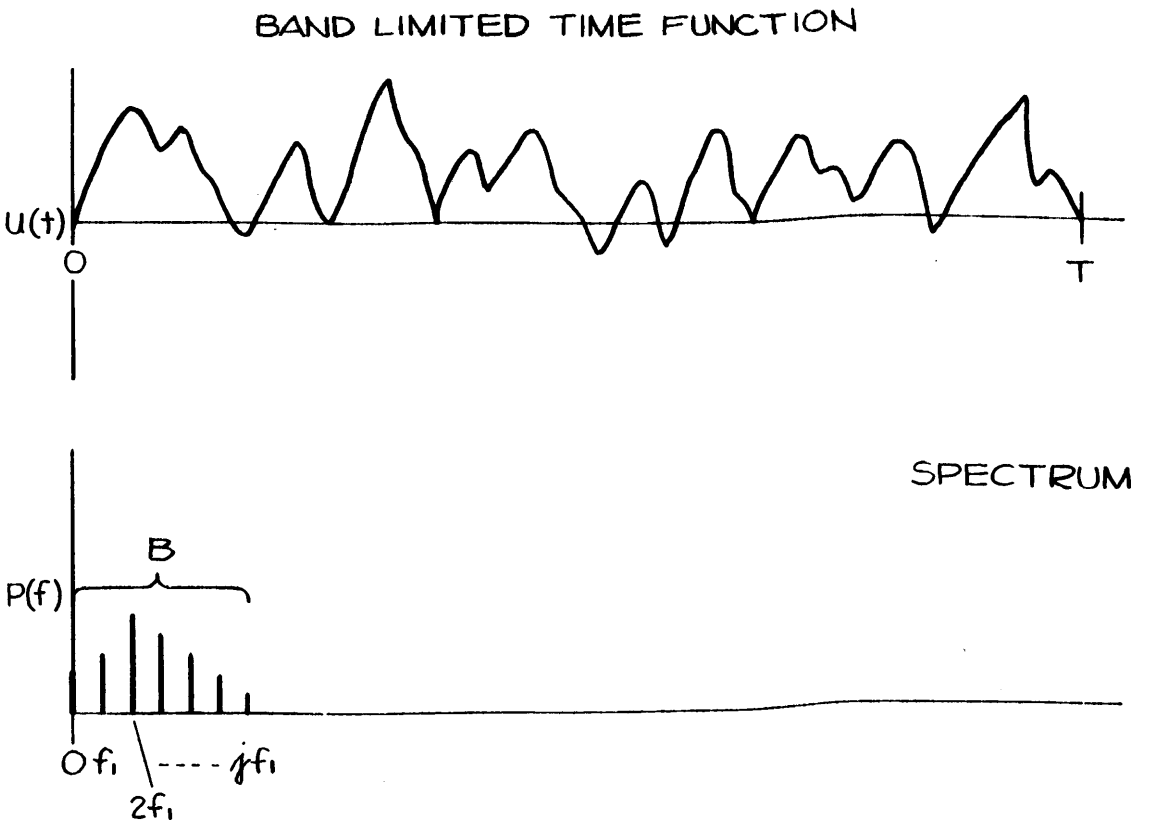
This theorem may be applied to either wide-band or narrow-band signals. The further result on narrow-band signals described by Woodward states that the samples obtained according to Theorem I

[1] P. M. Woodward, Probability and Information Theory, with Applications to Radar, pp. 31-37, Pergamon Press (1953).



$$u(t) = \sum_{k=h}^{k+q} (a_h \cos 2\pi h f_1 t + b_h \sin 2\pi h f_1 t)$$

Fig.6-2-FOURIER REPRESENTATION OF A NARROW BAND TIME FUNCTION



$$u(t) = \frac{a_0}{2} + \sum_{h=1}^{\infty} (a_h \cos 2\pi h f_1 t + b_h \sin 2\pi h f_1 t)$$

FIG. 6-3 - FOURIER REPRESENTATION OF A WIDE BAND TIME FUNCTION

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-15 | 5-21-65 BMB-ES

may be represented in terms of $B + \frac{\epsilon}{2}$ simultaneous amplitude and phase samples.

Theorem II:

> A narrow-band time function is completely specified by $B + \frac{\epsilon}{2}$ simultaneous amplitude and phase samples per second ($\epsilon > 0$).

These theorems may be referred to as the optimum sampling theorems. They specify the number of independent samples available per unit time and the maximum number of samples which are required if digital techniques are to be used in processing the function.

The actual sampling rate employed in practical processing depends upon the method which is to be used in reconstructing the time function and the functions derived from it in the computations. When sampling is done at a rate just satisfying Theorems I and II, reconstruction of u(t) from the samples may be accomplished by means of the terminated Fourier Series.

If sampling according to the relaxed rule to be quoted below is carried out, reconstruction by simple filter techniques is possible. With this technique, the relaxed or "practical" sampling theorem will be made plausible by describing the spectrum of the time function to be sampled and the spectrum of the sampled function. The capability of reconstructing u(t) by filter techniques will be clearly evident.

The sampling process may be considered equivalent to the product of u(t), the function to be sampled, and the function s(t), the sampling function. Figure 6-4 shows a time function u(t) and a sampling function s(t). The simplest sampling function is a series of equally spaced spikes. The spikes are τ apart. The product of s(t) and u(t) will have a value only at the position of the spikes. The product function will be the sampled version of u(t) and will be a series of spikes whose heights are

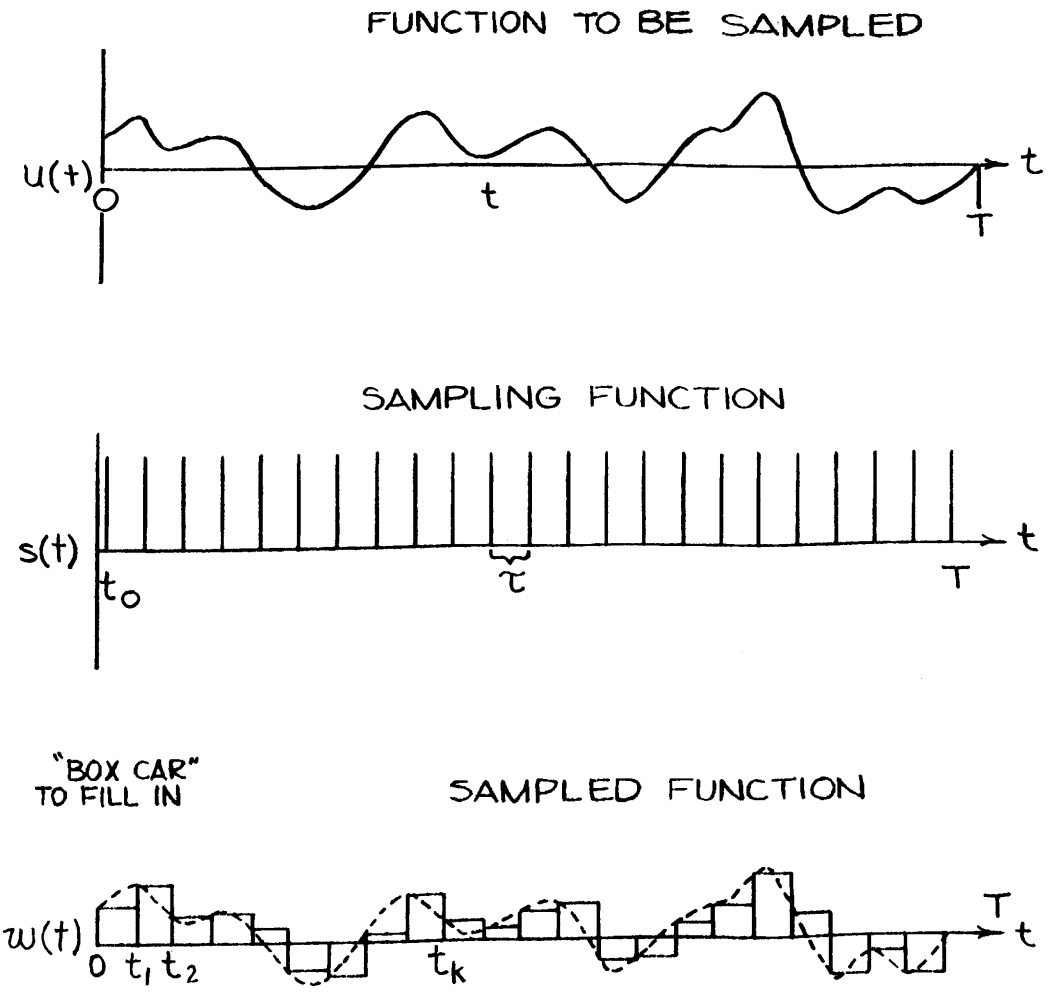


FIG. 6-4 SAMPLING AS A PRODUCT OF u(t) AND THE SAMPLING FUNCTION s(t).

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-16 | 5-21-65 BMB-ES

proportional to the values of u(t) at the sampling times. The sampling is completed by "boxcarring" the "spike" function. The sampled function then takes the form of the sequence of steps.

The spectra of u(t) and s(t) are shown in Fig. 6-5. The illustration is given for a narrow-band signal. The spectrum of u(t) is centered at f_o and has a width denoted by B. The spectrum of the sampling function is a series of spikes spaced along the frequency axis at intervals $f = \frac{1}{\tau}$, where τ is the sampling period.

The spectrum of the product of u(t) and s(t) is easily found from the identity

$$(a \cos 2\pi f_a t)(b \cos 2\pi f_b t) =$$

$$\frac{ab}{2} [\cos 2\pi(f_a + f_b)t + \cos 2\pi(f_a - f_b)t] .$$

For each frequency f_a present in u(t) and for each frequency f_b present in s(t), the frequencies $(f_a + f_b)$ and $(f_a - f_b)$ appear in the spectrum of the product. The lowest frequency present in the sampling function is $f = 0$; hence $f_o - 0$ and $f_o + 0$ both appear at f_o. The other frequencies in the bandwidth B transfer in the same way to a region around f_o in the spectrum of the sampled function. The next lowest frequency in the sampling function is $f_s = \frac{1}{\tau}$. This frequency will give rise to band centered at $(f_s - f_o)$ and another at $(f_s + f_o)$. The next lowest frequency $2f_s$ results in a band centered at $(2f_s - f_o)$ and at $(2f_s + f_o)$ (not shown) to the right of the edge of the figure.

The two versions of u(t), the original one and the sampled one, differ in that a number of higher spurious frequencies are now present in the spectrum of the sampled version. Recovery of the original u(t) from the sampled u(t) can be

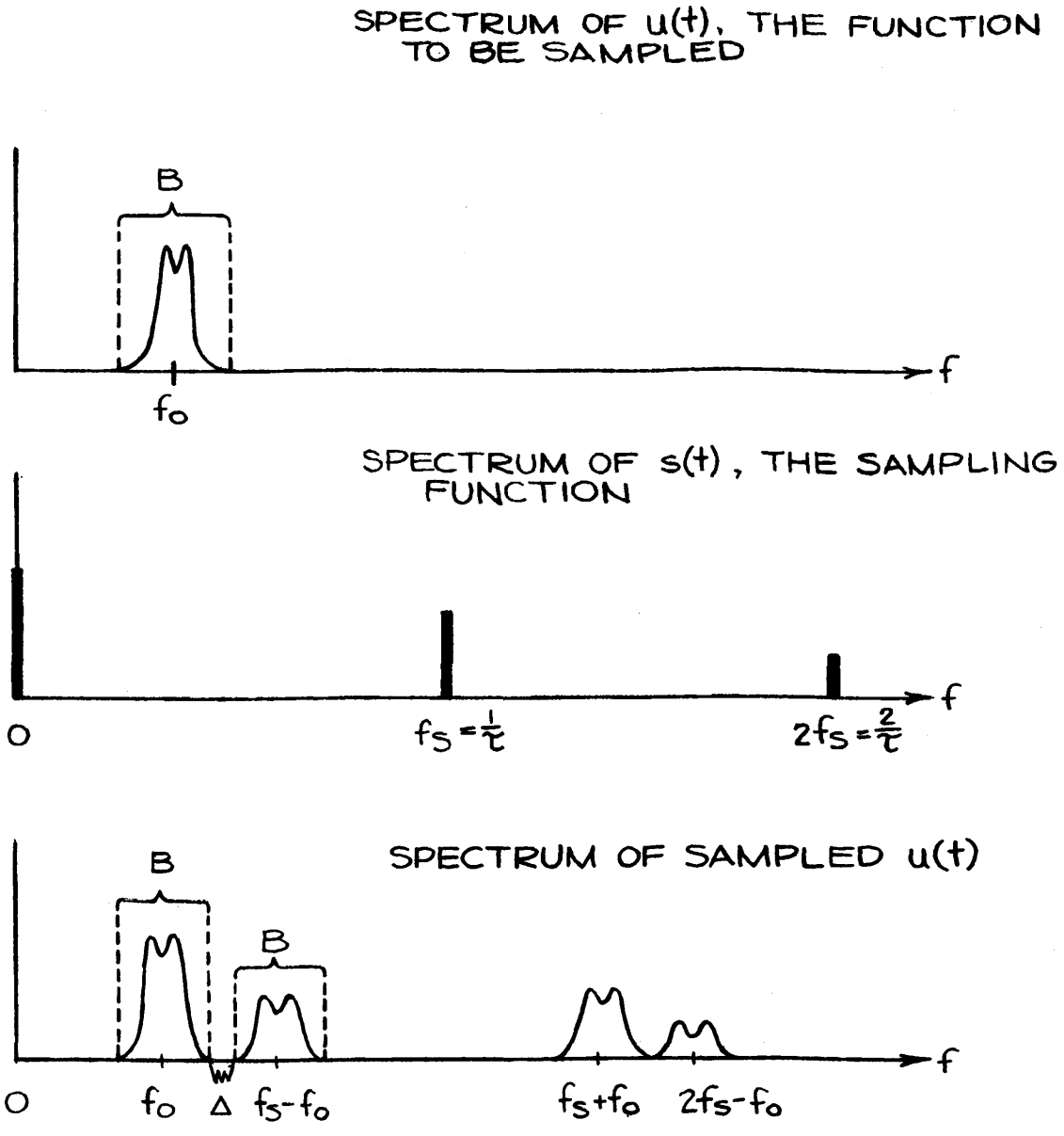


FIG. 6-5 SPECTRAL DESCRIPTION OF SAMPLED FUNCTION

accomplished provided none of the spurious frequencies fall in the band centered at f_o. This situation exists in the diagram; the nearest spurious band is separated from the band at f_o by an amount Δ. Δ will be greater than zero if

$$\Delta = (f_s - f_o - \tfrac{B}{2}) - (f_o + \tfrac{B}{2}) > 0. \qquad (6\text{-}10)$$

This condition is satisfied if f_s, the sampling frequency, satisfies the inequality,

$$f_s > 2(f_o + \tfrac{B}{2}). \qquad (6\text{-}11)$$

It will be noted that $(f_o + \frac{B}{2})$ is the highest frequency in the spectrum of u(t).

In order to reproduce u(t), the sampled function is played back through a narrow-band filter whose upper cut-off frequency is placed in the gap between the bands centered at f_o and $(f_s - f_o)$. If sufficiently steep skirted, this filter will remove a large portion of the power at the spurious spectral components and reconstruct u(t) to a very good approximation. Clearly this job is easiest to do when the gap Δ is made larger. With commercially available, adjustable three-stage filters, the smallest useful value of f_s is any value in excess of

$$f_s \sim 3(f_o + \tfrac{B}{2}) \qquad (6\text{-}12)$$

This sampling rate provides a gap between the spectral components of u(t) and the spurious components of the sampled u(t) in excess of $3(f_o + B/2)$.

When the spectrum of u(t) extends to zero frequency, the upper frequency in u(t) is B. The required f_s is

$$f_s > 2B\ , \qquad (6\text{-}13)$$

the rate required by the ideal sampling theorem. If the sampled function is to be reconverted to analog form by means of filters, a suitable value of f_s is

$$f_s \sim 3B \ . \qquad (6\text{-}14)$$

The value of f_s actually used can be any size greater than three times the highest frequency in u(t), but the expense of computing increases with the sampling rate (as the square of the sampling rate for correlation). It follows that extensive signal processing programs need to use the smallest feasible value of f_s.

The ideas presented concerning the "practical" sampling techniques indicate that it is possible to state a "Practical Sampling Theorem" which will provide enough equally spaced samples to allow reconstruction of functions to a good approximation by simple filter techniques.

Practical Sampling Theorem:

> Sampling at a rate corresponding to 2k times the highest frequency present in a time function will permit reconstruction of the function by simple n-stage filter techniques with spurious components more than 6kn dB below the reconstructed function.

The processing which might be carried on with a computer does not obviously make use of filter techniques in determining its answers. One illustration will serve to show that sampling before squaring does not give a result which is different from squaring before sampling. Figure 6-6 shows a waveform to be processed. Two alternate schemes are shown for

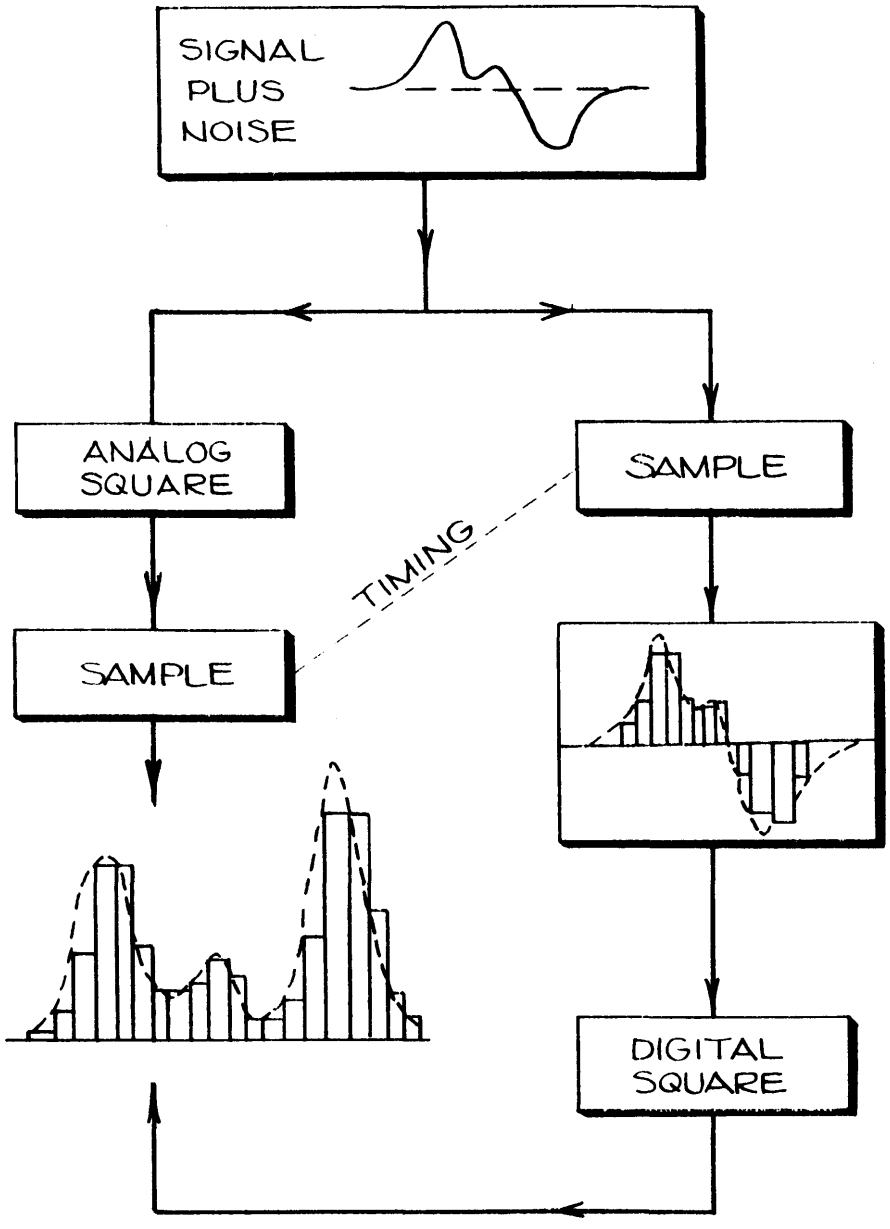


FIG. 6-6 ORDER OF SAMPLING AND SQUARING

TRACOR, INC. | DWG. 771-18
AUSTIN, TEXAS | 5-24-65 BMB-ES

obtaining a squared version of a sampled waveform. Along the left channel the waveform is squared and then sampled. In the process the points on the dotted envelope are converted to points on the rectangular envelope. The left-most points in the rectangles of the sampled version are identical to the dotted envelope at those points. In the channel on the right, if the points of the waveform are first sampled (in synchronism with the samples in the left channel) the bipolar sampled version will be obtained. The left-most corner of each rectangle is identically equal to the value of the original waveform at the same point. Digital squaring will make the left-most corner of each rectangle identical to the square of the original waveform at the same point. Clearly both these processes give the same sampled waveform. Since the left channel, analog squaring followed by sampling, permits reconstruction of the squared waveform according to the sampling rules already laid out, sampling before squaring, which gives the same squared sampled function, also permits reconstruction of the complete squared waveform.

A by-product of the discussion of sampling which can be obtained from the finite Fourier series (Eq. 6-3)) used to represent an arbitrary time function u(t) is the concept of frequency resolution. In the bandwidth B of u(t) there are only

$$kf_1,\ (k+1)f_1,\ (k+2)f_1,\ \ldots,\ (k+j)f_1,$$

a total of $j + 1$ frequencies. This fact suggests that attempts to determine the presence of power at frequencies between two of the specified ones, say by means of narrow-band filters will provide only weighted averages of the power at the frequencies present. No further indication of the basis of this property will be given here, but a general frequency resolution theorem will be quoted without proof. Further discussion of this

property in special cases will be given in a later section.

Frequency Resolution Theorem:

> In a time function u(t) having duration T, the power at frequency f_1 and at frequency f_2 can have statistical independence only when $f_2 - f_1 \geq \frac{1}{T}$.

A similar statement can be made about power when the time function is a narrow-band function because power depends only upon the waveform amplitude. Since $B + \frac{\epsilon}{2}$ ($\epsilon > 0$) amplitude samples per second suffice to describe the power in the time function, a new power sample can occur only each $1/(B + \epsilon/2)$ seconds. This fact leads to a time resolution for power changes.

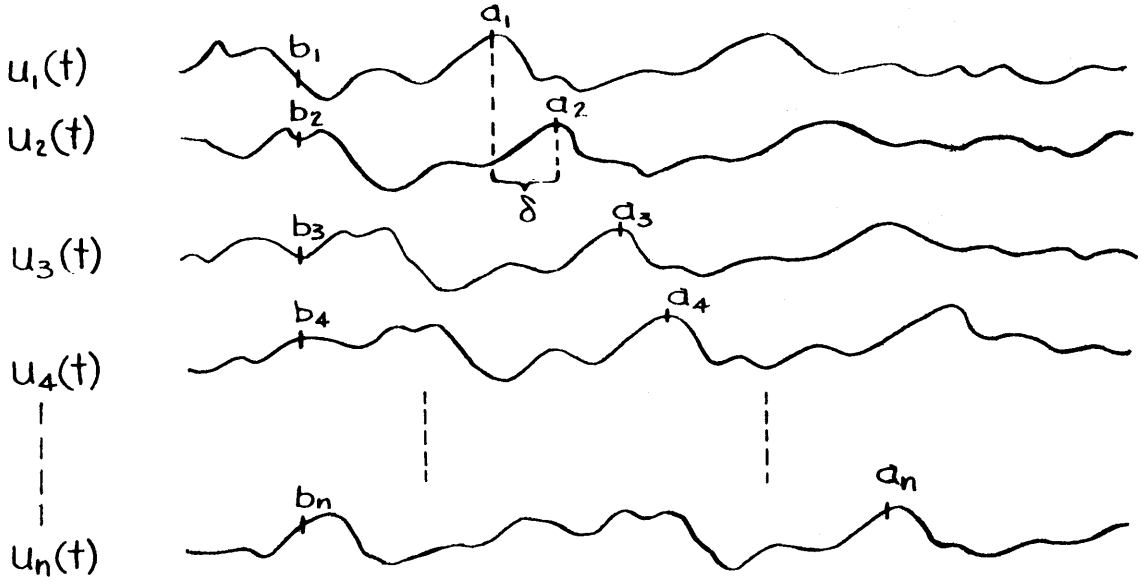
Time Resolution Theorem:

> In a time function u(t) having bandwidth B, the power at time t_1 and the power at time t_2 can have statistical independence only when $t_2 - t_1 \geq \frac{1}{B}$.

6.1.2 Additional Theorems

Quite often in the study of signal processing the sum of a number of quantities must be determined. When those quantities are samples of a time function taken in sequence or when they are simultaneous samples in a number of channels, the sum exhibits two kinds of limiting behavior, depending upon whether the samples are coherent or incoherent. The incoherent quantities generally arise from the random variation of background while coherent quantities result when the samples are taken from periodic time functions or two or more channels containing signals, identical except for relative time shift.

Two examples are illustrated by Fig. 6-7, which shows a number of time functions representing the outputs $u_i(t)$



COHERENT

$$A_n = \sum_{i=1}^{n} a_i = n a_1$$

$$\overline{A_n^2} = n^2 \overline{a^2}$$

INCOHERENT WITH ZERO MEAN

$$\overline{b_i} = \frac{1}{n} \sum_{i=1}^{n} b_i = 0$$

$$\overline{b_i^2} = \frac{1}{n} \sum b_i^2$$

$$B_n^2 = \sum b_i^2 = n\overline{b_i^2}$$

FIG 6-7 ADDITION OF SAMPLES FROM A SET OF TIME FUNCTIONS

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-19 | 5-24-65 BMB-ES

of a number of hydrophones in an array. For this example the waveform originates at a noise source off the axis of the receiving beam so that the same waveform emerges from each hydrophone with a progressive delay δ from one element to the next. The two sets of time function samples are marked: the a_i which are corresponding samples if suitable delays, δ, are introduced from one hydrophone to the next and the b_i which are simultaneous samples at the hydrophone outputs. The b_i are statistically independent by virtue of the nature of the noise source. The a_i are identical samples because of the delay insertion. The sum

$$A_n = \sum_{i=1}^{n} a_i = na_1 \ , \tag{6-15}$$

of the coherent samples is straightforward. The square A_n^2 is the instantaneous power which would be obtained from the array at the instant the delayed a_i emerged from the channels:

$$A_n^2 = n^2 a_1^2 \ . \tag{6-16}$$

The average power from the transducer under coherent addition is $\overline{P}_c$ given by

$$\overline{P}_c = \overline{A_n^2} = n^2 \overline{a^2} \ . \tag{6-17}$$

In this expression $\overline{a^2}$ represents the average square of all the samples from a single hydrophone; clearly $\overline{a^2}$ can be written $\overline{u^2(t)}$. The average power $\overline{P}_c$ from the array which results from coherent summing can now be written

$$\overline{P}_c = n^2 \overline{u^2(t)} \ . \tag{6-18}$$

The incoherent sum arises if the delays are not inserted and the simultaneous outputs are the b_i. The sum in this case is the sum of a number of statistically independent numbers all selected from ~~similar~~ similar waveform distributions. It is well known that the mean μ and standard deviation σ of the sum of n quantities selected from a distribution with mean μ_1 and standard deviation σ_1 are given by

$$\mu = n\mu_1 \tag{6-19}$$

$$\sigma^2 = n\sigma_1^2 \ . \tag{6-20}$$

In these expressions μ_1 and σ_1 are defined by

$$\mu_1 = \frac{1}{n}\sum_{i=1}^{n} b_i \qquad \sigma_1^2 = \frac{1}{n}\sum_{i=1}^{n} (b_i - \mu_1)^2 \tag{6-21}$$

See for example Woodward's[2] concise discussion. Any standard, elementary treatment of statistics will do as well. For the situation assumed for this example, the same noise function u(t) emerges from each hydrophone but with a delay from one to the next. The sequence of the b_i represents a sampled version of u(t) so that the time averages of u(t) are equivalent to the corresponding averages of the b_i. In this case μ_1 and σ_1 can be written

$$\mu_1 = \overline{u(t)} \qquad \sigma_1^2 = \overline{[u(t) - \mu_1]^2} = \overline{u^2(t)} - \overline{u(t)^2}. \tag{6-22}$$

---

[2] P. M. Woodward, Probability and Information Theory, p. 9 Pergamon Press (1953).

This equivalence of the time averages and the so-called ensemble averages does not hold in general. If the transducer outputs meet the requirements of ergodicity, such is the case, but there is seldom a practical situation where enough information is known to confirm ergodicity. It is usually assumed that time averages and ensemble averages can be equated.

The previous paragraph has been inserted in order to point out that the equivalence of time averages and ensemble averages cannot be taken for granted. The discussion was entirely unnecessary for this example because the hydrophone outputs have been assumed to be time delayed replicas of the noise waveform. In this example the mean μ_1 of the output of a single hydrophone is assumed to be zero so that the standard deviation of the sum of the b_i is, according to Eqs. (6-20) and (6-22)

$$\sigma^2 = n\sigma_1^2 = n\,\overline{u^2(t)} \quad . \tag{6-23}$$

This is, of course, the power output $\overline{P_I}$ of the array when the noise source is off-axis, i.e., where the hydrophone outputs are summed incoherently.

$$\overline{P_I} = n\,\overline{u^2(t)} \quad . \tag{6-24}$$

It is interesting to compare the power output when the noise source is on the array axis and when it is off-axis. This ratio is the ratio expected between the result of coherent summing and of incoherent summing. The ratio results when Eq. (6-17) is divided by Eq. (6-24):

$$\frac{\overline{P_c}}{\overline{P_I}} = \frac{n^2\,\overline{u^2(t)}}{n\,\overline{u^2(t)}} = n \quad . \tag{6-25}$$

The only assumptions used in arriving at this result are (1) the

same time function emerges from each hydrophone, (2) when the source is on the array axis the functions are in time registry, and (3) when the noise source is off-axis the values of the time function from one hydrophone to the next are statistically independent. Not all these assumptions are met in practice with an actual transducer. In the first place there are other local sources of transducer excitation than distant sources so that the output of all transducers is not the same; however, to a good approximation the part of the output which arises from a single distant source is the same for each hydrophone. In the second place the transducers in an array are usually about one-half wavelength apart so that the outputs of adjacent transducers to off-axis sources are not completely statistically independent. In a practical array the ratio P_C/P_I which can be achieved is within 3 dB of the ratio given by Eq. (6-25). This means that the array behaves about as if n/2 of its elements furnish statistically independent samples. This ratio is the source of array gain or directivity index by which the sonar is able to discriminate against noise sources not on its axis.

6.1.2.1 The Detector Averager. Suppose a waveform with bandwidth B is introduced into a detector followed by an RC filter with time constant τ. This waveform is to be a random fluctuation. According to the sampling theorem discussed in Section 6.1.1, there will be 2B statistically independent input waveform samples per second. At the output there will be $2/\tau$ independent samples per second. This result is obtained by making use of the fact that the bandwidth at the output of the RC filter is about $1/\tau$. The number of statistically independent input samples per output sample is

$$n = 2B/(2/\tau) = B\tau \ . \qquad (6\text{-}26)$$

The mean μ_o of the waveform at the averaged output is

$$\mu_o = n\,\mu_1 = B\tau\,\mu_1 \quad , \tag{6-27}$$

where μ_1 is the mean of the absolute value of the input waveform. The standard deviation σ_o at the output is given by

$$\sigma_o^2 = n\,\sigma_1^2 = B\tau\,\sigma_1^2 \quad , \tag{6-28}$$

where σ_1 is the standard deviation of the input waveform. It follows that

$$\frac{\mu_o}{\sigma_o} = \sqrt{n}\,\frac{\mu_1}{\sigma_1} = \sqrt{B\tau}\,\frac{\mu_1}{\sigma_1} \quad . \tag{6-29}$$

Since μ_1 and σ_1 are presumably fixed for a given waveform source, the ratio of μ_o, the average output, to σ_o, the root mean square deviation from the mean (the wiggliness) of the output waveform, increases as the square root of $B\tau$. This means that the rms error made in measuring μ_o decreases by $\sqrt{B\tau}$ or by the square root of the number of independent input samples added to produce a single output sample. A typical fractional spread at the output,

$$\frac{\mu_o \pm \sigma_o}{\mu_o} = 1 \pm \frac{1}{\sqrt{n}}\,\frac{\sigma_1}{\mu_1} \quad ,$$

when converted to dB, becomes

$$\text{dB spread} = 20 \log\left[\frac{\mu_o \pm \sigma_o}{\mu_o}\right] \sim \pm 8.7\,\frac{1}{\sqrt{n}}\,\frac{\sigma_1}{\mu_1} \quad . \tag{6-20}$$

If σ_1 and μ_1 are the same order of magnitude, the result $8.7/\sqrt{n}$ is an estimate of the error made in estimating the average value of the output waveform by simply reading the voltage at the output of the averager.

Table I

dB Spread in which 70% of Measurements Will Lie

$n = B\tau$	1	3	10	30	100
Approx. Spread (dB)	±6	±4	±3	±1.5	±1

It is seen from the tabulated values that fluctuations in the output of the order of ±3 dB are common when the $B\tau$ product is as high as 10. Quite a large number of independent samples are required to produce an output with fluctuations of the order of ±1 dB.

6.1.2.2 The Measurement of Average Power. This example will amount to an extension of the previous one. Suppose it is desired to determine the average power of a time function u(t). If the function has bandwidth B, 2B statistically independent samples may be obtained each second. If it is desired to determine the average power of u(t) with a specified precision from the independent samples, a number η will be required. The computation will be

$$P_\eta = \frac{1}{\eta} \sum_{i=1}^{\eta} u^2(t_i). \qquad (6\text{-}31)$$

The number η must be chosen so that the value P_η, the mean of $u^2(t)$, and the standard deviation of P_η are related like μ_o and σ_o of the last example.

When P_η is to be an estimate of the average power of the entire waveform accurate to about 1 dB, then η must be ≥ 100 according to the previous table. The length T of time function required to be sampled is obtained from the expression

$$BT \geq 100 \qquad T \geq \frac{100}{B}. \qquad (6\text{-}32)$$

For a system admitting a waveform with a bandwidth B of 200 c/s, it is therefore necessary to sample at least 0.5 seconds of the waveform to obtain a result which 70% of the time lies within 1 dB of the average power of the entire waveform. These statements are of course based upon the assumption that the nature of the source producing the waveform is not altered either during, before, or after the sampling takes place.

Blackman and Tukey[3] discuss this problem from the point of view of χ^2 distributions being representative of the probabilities encountered. In their table they provide an estimate of the number of independent samples or degrees of freedom required for a few tabulated values of dB spread with the requirement that the result fall within the spread a tabulated fraction of the time. The Table I values included in this discussion are in approximate agreement with their tabulated values for results falling within the spread about 70% of the time.

6.1.3 The Measurement of Power Spectra

In the section dealing with sampling it was shown that frequency resolution depended upon the length of the time sample. This result is very important in the determination of power spectra. In this discussion a sampled version of a noise function will be obtained by sampling u(t) over a time $0 \leq t \leq T$. The power spectrum of u(t) will be estimated by squaring the absolute value of the Fourier transform F(f) of the sampled section of u(t) given by

$$F(f) = \int_{-\infty}^{\infty} u(t)\, e^{-i2\pi ft}\, dt \ . \qquad (6\text{-}33)$$

[3] R. B. Blackman and J. W. Tukey, The Measurement of Power Spectra, p. 23, Dover Publications, New York (1959).

No samples have been taken outside the interval $0 \leq t \leq T$ so that the integration needs to be performed only over this interval. Since values of F(f) corresponding to two frequencies f_1 and f_2 closer together than 1/T are not statistically independent, the transform need be computed only at a set of frequencies

$$f_1,\ f_1 + 1/T,\ f_1 + 2/T,\ \ldots,\ f_1 + k/T\ , \qquad (6\text{-}34)$$

k + 1 frequencies being required to cover the band of interest.

Consider the result obtained at the ith frequency,

$$f_i = f_1 + \frac{i - 1}{T} \qquad (6\text{-}35)$$

A value $F(f_i)$ is determined. The question may be asked, "How well does $F^2(f_i)$ represent the spectral density of the entire u(t), including the unsampled portion?" (It is of course assumed that the source of u(t) does not change character before or after sampling.) Since 2BT+2 independent samples are available and k/T is the bandwidth of interest (k/T = B) it follows that k = BT, hence the number of degrees of freedom (statistically independent samples available) per frequency is about

$$\frac{2BT+2}{k+1} = \frac{2BT+2}{BT+1} = 2\ . \qquad (6\text{-}36)$$

This is not very many. A single time function of length T can provide the power density at f_i with an accuracy corresponding to only 2 statistically independent samples. According to Table I the expected spread in the result is about ±5 dB.

How can this result be improved? The obvious and very satisfactory way is to sample several sections (a,b,c, etc.) of u(t) of length T. Several values of the transform $F(f_i)$

$$F_a(f_i),\ F_b(f_i),\ F_c(f_i),\ \text{etc.}$$

will be obtained from the several sections. When these are squared and averaged, the total number of statistically independent samples will be twice the number of sections employed. Thus with five sections of length T of the time function 10 degrees of freedom will be obtained and $F^2(f_i)$ averaged over the five sections will be characterized by an expected spread of ±3 dB.

Before leaving the subject of power spectra, the spectra of two useful sonar pulses should be described. The spectra can be determined using Eq. (6-33); the Fourier transform is computed and the spectral density is then $F^2(f)$.

The first example is the sinusoidal pulse of length T, (see Fig. 6-8) defined by

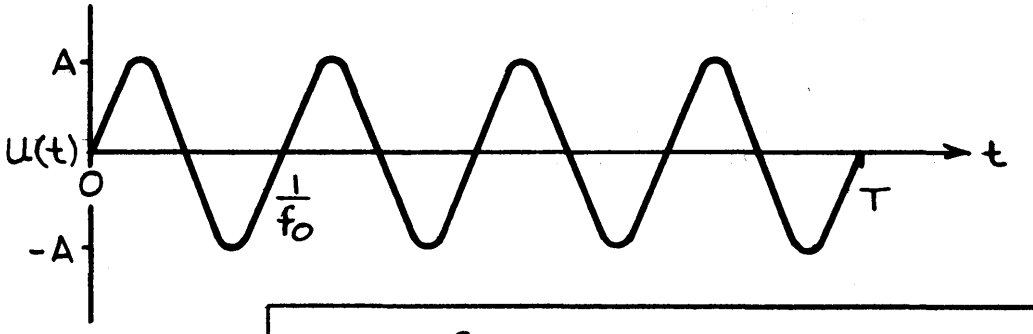
$$u(t) = \begin{cases} A \sin 2\pi f_o t & -T/2 \le t \le T/2 \\ 0 & |t| > T/2 \end{cases} . \qquad (6\text{-}37)$$

The Fourier transform is, according to Eq. (6-3),

$$F(f) = A \int_{-T/2}^{T/2} \sin 2\pi f_o t \; e^{-i2\pi ft}\, dt . \qquad (6\text{-}38)$$

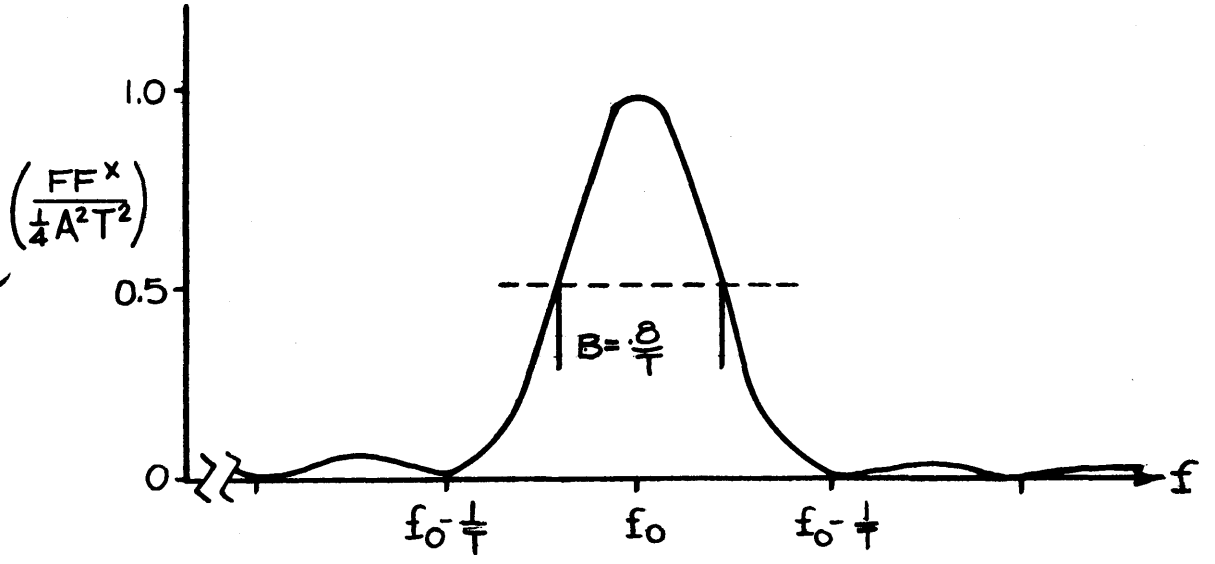
When $\sin 2\pi f_o t$ is written $(e^{i2\pi f_o t} - e^{-i2\pi f_o t})/2i$, the integral takes the form

$$F(f) = A/2i \int_{-T/2}^{T/2} [e^{i2\pi(f_o - f)t} - e^{-i2\pi(f_o + f)t}]\, dt. \qquad (6\text{-}39)$$



$$u(t) = \begin{cases} A \sin 2\pi f_o t & 0 \le t \le T \\ 0 & t < 0 \text{ OR } t > T \end{cases}$$

(a) TIME FUNCTION



$$F(f)\,F^x(f) = \frac{A^2T^2}{4}\left[\frac{\sin T(f-f_o)T}{\pi(f-f_o)T)}\right]^2$$

(b) SPECTRAL DENSITY

FIG. 6-8 THE SINUSOIDAL PULSE

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-42 | 6-24-65 BMB-BS

After integration $F^2(f)$ has the form

$$|F(f)|^2 = \left[\frac{A^2T^2}{4}\right]\left[\frac{\sin \pi(f_o - f)T}{\pi(f_o - f)T} + \frac{\sin \pi(f_o + f)T}{\pi(f_o + f)T}\right]^2 . \quad (6\text{-}40)$$

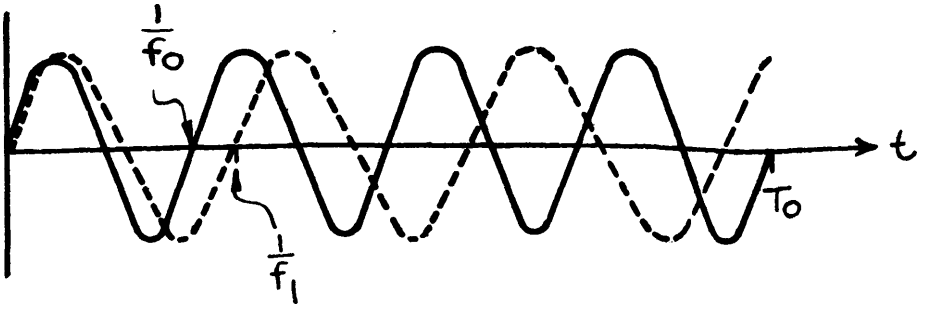
The part of the spectrum of the pulse around $f = f_o$ is shown in Fig. 6-8. There is a similar portion around $f = -f_o$. The spectrum is localized near these frequencies, the first zero on either side of the main peaks being separated from $\pm f_o$ by an amount 1/T. The half power points on the main peaks are separated by about 0.8/T. About 90% of the power is within a band of width 1/T centered on $\pm f_o$. These results are in approximate agreement with the usual assumption that the bandwidth B of the sinusoidal pulse is

$$B = \frac{1}{T} . \quad (6\text{-}41)$$

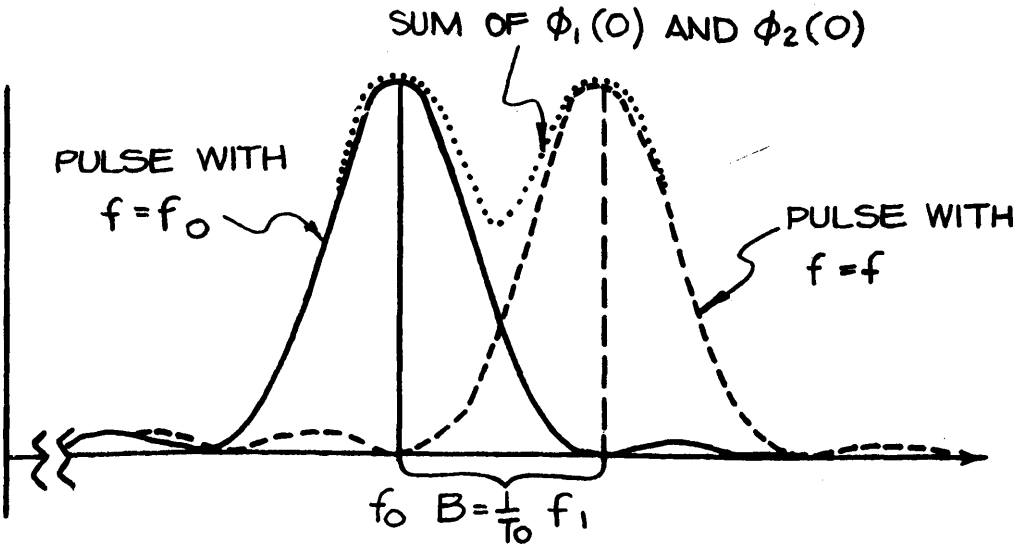
A long sinusoidal pulse has a narrow spectrum while a short one has a broad spectrum.

An illustration of the frequency resolution theorem is easily discussed in terms of the spectrum shown in Fig. 6-8 and 6-9. Suppose there are two frequencies f_1 and f_o present in the function u(t). The observed combined spectrum will consist of a broad peak or perhaps two peaks. If the two peaks are separated so that the peak of the f_1 spectrum falls at the first zero of the f_o spectrum, the sum plot will have a 1 dB dip between the two peaks. If the separation of the individual peaks is reduced reduced 11% the minimum has disappeared and a single broad peak is observed. If the separation of the individual peaks is increased the peaks are well resolved.

The spectral behavior described in the preceding paragraph represents the power outputs of a large number of frequency-contiguous, narrow-band filters plotted against the



(a) TWO SINUSOIDAL PULSES



(b) SPECTRAL DENSITIES SHOWN RESOLVED IF
$(f_1 - f_0) \geq \frac{1}{T_0}$

FIG.6-9 FREQUENCY RESOLUTION OF SINUSOIDAL PULSES

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-43 | 6-24-65 BMB-ES

center frequency of the filters. This result is of great practical importance in determining the presence or absence of more than one frequency in a time function u(t).

According to Eq. (6-40), the first zero near the peak at f_o is separated from f_o by $\frac{1}{T}$. Resolution occurs then if

$$f_1 - f_o \geq \frac{1}{T} ,$$

exactly as claimed in the frequency resolution theorem described in the section concerning sampling methods.

One further interesting result can be mentioned: The total number of cycles N_o in the pulse with frequency f_o is

$$N_o = f_o T ,$$

while the number N_1 in the pulse with frequency f_1 is

$$N_1 = f_1 T .$$

From these it is clear that

$$f_1 - f_o = \frac{N_1 - N_o}{T} \geq \frac{1}{T} ,$$

whence

$$N_1 - N_o \geq 1 .$$

This is another form of the frequency resolution theorem applicable to sinusoidal pulses. It states that two pulses of the same duration can just be distinguished as having

different frequencies if one of the pulses contains one more cycle than the other.

The second example is the FM slide or FM sweep. This pulse has the form

$$u(t) = \begin{cases} A \sin 2\pi(f_o + \frac{at}{2})t , & \frac{-T}{2} \le t \le \frac{T}{2} \\ 0 & |t| \ge T/2 \end{cases} \qquad (6\text{-}42)$$

In this pulse the frequency shifts from

$$f_o - \frac{aT}{2} \text{ to } f_o + \frac{aT}{2}$$

during transmission, a frequency shift δ_f amounting to

$$\delta_f = aT \quad . \qquad (6\text{-}43)$$

The quantity a is the frequency shift per second.

This pulse can also be inserted into Eq. (6-33). The integration is straightforward but a little more troublesome so the result will be quoted without showing the detail of the calculation. The spectrum of the 200 ms, 100 c/s FM slide is shown in Fig. 6-10. The spectrum of the 200 ms sinusoidal pulse (CW) is shown (not to the same scale - the areas under the two pulses are equal if they are plotted to the same scale) for comparison. It will be noted that if the areas under the two plots are equal that the FM slide has a low, flat spectrum while the CW pulse has a taller, narrow spectrum.

6.1.4 The Correlation Function

The properties of the correlation function simplify the description of optimum filtering techniques. The more useful

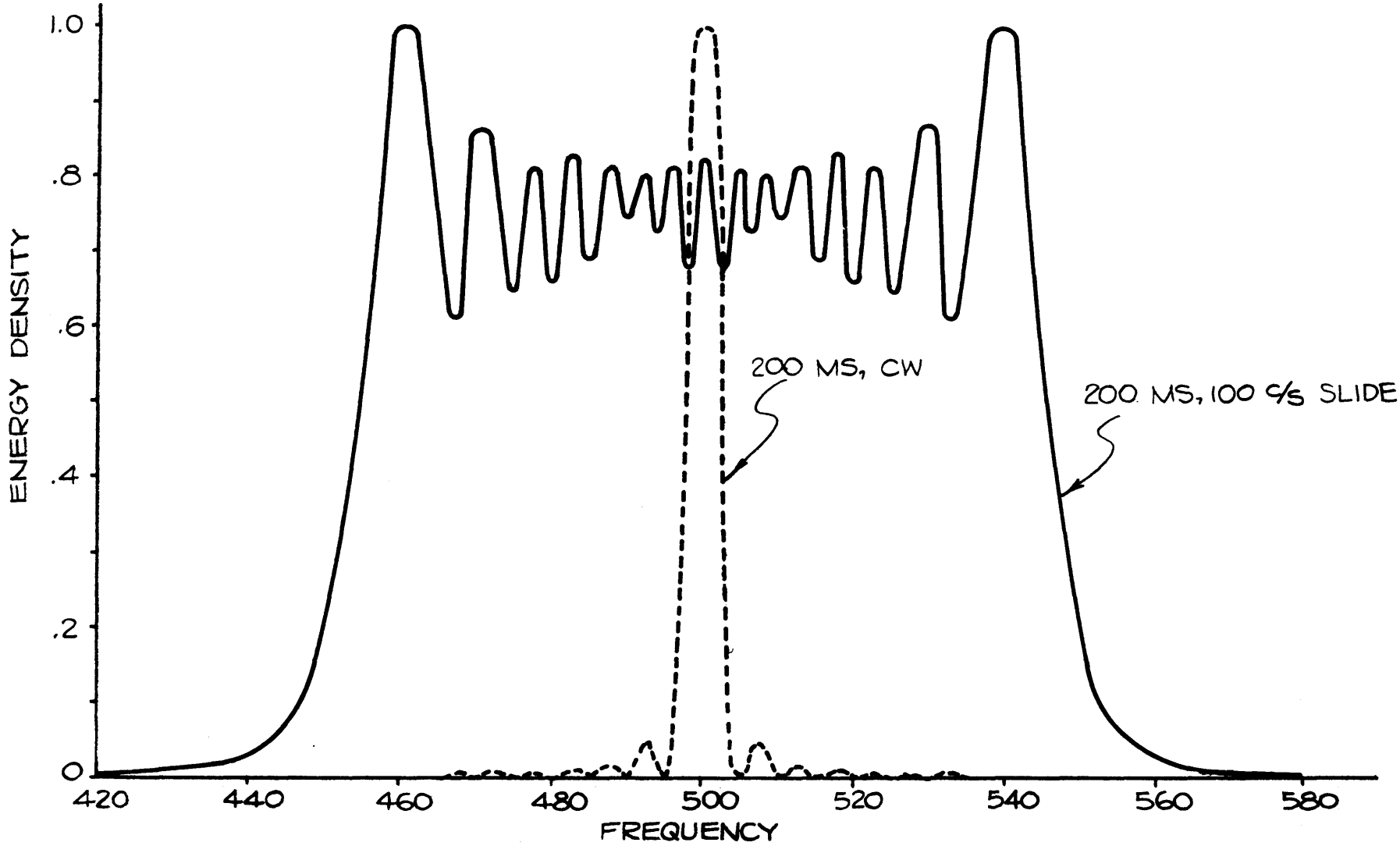


FIG. 6-10 - SPECTRA OF TYPICAL PULSES

properties will be described in this section and their application will be discussed in the next section. The correlation function $\varphi_{uw}(\delta)$ is a property of the product of two time functions u(t) and w(t). It is defined as

$$\varphi_{uw}(\delta) = \int_{-\infty}^{\infty} u(t)\, w(t + \delta)\, dt \ ; \qquad (6\text{-}44)$$

δ is a time shift between the functions u(t) and w(t). In general, u(t) and w(t) are not functions of infinite duration but exist only between well defined time limits. For the purpose of this discussion, it will be assumed that u(t) starts at t = 0 and lasts until T, while the other w(t) starts at T_1 and lasts until T_2. Furthermore, it will be assumed that both functions carry unit energy. This means that

$$\int_{0}^{T} u^2(t)\, dt = 1 \ ,$$

$$\int_{T_1}^{T_2} w^2(t)\, dt = 1, \qquad (6\text{-}45)$$

While this may seem to be a restriction, it is not serious. If it is desired to represent a function h(t) which is similar to w(t) except in the ratio A of their sizes it is merely necessary to write

$$h(t) = A\, w(t) \ , \qquad (6\text{-}46)$$

so that any results obtained for w(t) will also apply for h(t) if they are suitably scaled by the factor A. It follows then that

$$\int_{T_1}^{T_2} h^2(t)\, dt = A^2 \int_{T_1}^{T_2} w^2(t)\, dt = A^2 \,, \qquad (6\text{-}47)$$

for example.

The properties of the correlation function germane to signal processing are most easily discussed in terms of the sampled versions of u(t) and w(t). If the time between samples is τ taken in accordance with the sampling theorem, the functions are representable with $2BT + 2 = n$ and $2B(T_2 - T_1) + 2 = m$ samples.

$$\begin{aligned} u(t):&\ u_1, u_2, \ldots, u_i, \ldots u_n \\ w(t):&\ w_1, w_2, \ldots, w_i, \ldots w_m \,. \end{aligned} \qquad (6\text{-}48)$$

It will be assumed that $m > n$. With the sampled representation of u(t) and w(t), $\varphi_{uw}(\delta)$ takes the form,

$$\varphi_{uw}(j\tau) = \sum_{i=1}^{n} u_i\, w_{j+i} \,. \qquad (6\text{-}49)$$

In the sampled form δ can take values equal only to multiples of τ. In the sampled notation the conditions of Eqs. (6-45) become

$$\begin{aligned} \sum_{i=1}^{n} u_i^{\,2} &= 1 \,, \\ \sum_{i=1}^{m} w_i^{\,2} &= 1 \,, \end{aligned} \qquad (6\text{-}50)$$

and Eq. (6-47) becomes

$$\sum_{i=1}^{m} h_i^2 = A^2 \sum_{i=1}^{m} w_i^2 = A^2 \ . \qquad (6\text{-}51)$$

It will be noted that Eq. (6-49) resembles the "dot" or scalar product of two n-component vectors $\mathbf{u}$ and $\mathbf{w}$:

$$\mathbf{u} \cdot \mathbf{w} = \sum_{i=1}^{n} u_i \, w_{j+i} = |u| \ |w| \cos \zeta_j \ . \qquad (6\text{-}52)$$

in which ζ_j is the n-dimensional analog of the angle between $\mathbf{u}$ and $\mathbf{w}$. This reduces to

$$\varphi_{uw}(j\tau) = \sum_{i=1}^{n} u_i \, w_{j+i} = \cos \zeta_j \qquad (6\text{-}53)$$

because the magnitudes of $\mathbf{u}$ and $\mathbf{w}$ are each unity. It follows that $\varphi_{uw}(j\tau)$ has a maximum value 1 when u_i and w_j are parallel--when u_i and w_j are the same vector--when each component u_i of $\mathbf{u}$ is identical to its mate w_{j+i} of w.

This result can be expressed using the coherent addition theorem: If the u_i and w_{j+i} are equal, pair by pair,

$$\varphi_{uw}(j\tau) = \sum_{i=1}^{n} u_i w_{j+i} = \sum_{i=1}^{n} u_i^2 = n \, \overline{u_i^2} = \text{Energy in } u(t) \qquad (6\text{-}54)$$

In this case $\varphi_{uw}(j\tau) = \varphi_{uu}(j\tau)$ is called the autocorrelation function.

When the vectors are not equal, when for example, the u_i and corresponding w_{j+i} are uncorrelated random noise functions having zero mean, the mean over j of the sum of the products will be zero as well. The variance of the sum of the products will be the mean of $\sigma^2(j)$ given by

$$\sigma^2(j) = \frac{1}{n} \sum_{i=1}^{n} (u_i^2 \, w_{j+i}^2) \tag{6-55}$$

which on the average is $\sigma^2 = \sigma^2(j)$:

$$\sigma^2 = n \, \overline{u_i^2} \, \overline{w_j^2} = n(\overline{u_i^2})^2 \, , \tag{6-55a}$$

if $\overline{u_i^2}$ and $\overline{w_j^2}$ are about the same size as they will be if m and n are about the same. The correlation function of these two uncorrelated "vectors" averages about $\pm\sigma$

$$\varphi_{uw}(j\tau) \cong \pm\varphi \sim \frac{\text{Energy in } u(t)}{\sqrt{n}} \tag{6-56}$$

It follows that the autocorrelation function of u(t) evaluated at $\delta = 0$ divided by the correlation function of u(t) and w(t) averaged over δ is

$$\frac{\varphi_{uu}(0)}{\varphi_{uw}(\delta)} = \sqrt{n} \, , \tag{6-57}$$

where n is the number of independent samples in the shorter of the functions. This is equivalent to

$$\left[\frac{\varphi_{uu}(0)}{\varphi_{uw}(\delta)}\right]^2 = n = 2BT \, . \tag{6-58}$$

The square of the ratio of the autocorrelation of u(t) at zero

shift to the correlation for u(t) and a random background function w(t) is on the average 2BT.

An interesting set of functions for use in sonar is the function which is transmitted and the doppler shifted versions of the functions which might represent the echo from a moving, localized target. Suppose the transmitted function is $u(f_o, t)$ in which f_o represents the center frequency of the transmission and t represents time. The echo which returns from the moving target will be

$$k\, u(f, t - \delta),$$

an amplitude-scaled, time-delayed, and doppler-shifted replica of the transmitted pulse. It will be assumed that signal levels are very large so that noise can be ignored in this example.

$$\sum_{i=1}^{n} u^2(f_o, t_i) = 1 \tag{6-59}$$

and

$$k^2 \sum_{i=1}^{n} u^2(f, t_i - \delta) = k^2 . \tag{6-60}$$

Suppose a number of frequency shifted replicas of $u(f_o, t)$, are generated and sampled:

$u(f_1, t)$,

$u(f_2, t)$,

...

....

$u(f_n, t)$.

The frequency spacing of these functions is determined by the frequency resolution $\delta f = 1/\tau$ of the transmitted pulse. Suppose further that the correlation functions of the form

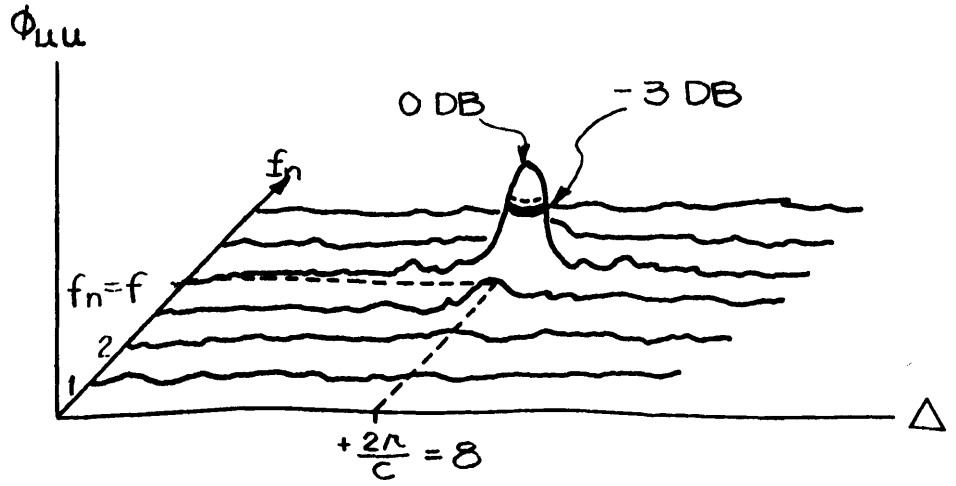
$$\varphi(f_n, f, \Delta) = \sum_{i=1}^{n} u(f_n, t_i - \Delta)\, ku(f, t_i - \delta) \qquad (6\text{-}61)$$

are computed. This is the correlation function of the amplitude-scaled (by k), time-delayed (by δ seconds), and doppler-shifted (to f) echo with one of the artificially produced, frequency-shifted (to f_n) replicas of the transmitted pulse, delayed by an amount Δ. For a particular f_n it has the appearance of one of the nearly parallel three-dimensional plots in the upper half of Fig. 6-11. As n increases, 1, 2, ..., a value of n is reached at which small peaks begin to appear near the delay Δ corresponding to the round trip travel time 2r/c to the target. The largest peak occurs for $f_n = f$ and $\Delta = \delta = 2r/c$. In this case the correlation function is

$$\varphi(f_n = f, f, \Delta = \delta) = k \sum_{i=1}^{n} u^2(f, t_i - \delta) = k\ ; \qquad (6\text{-}62)$$

the locally generated replica and the reflected replica are parallel vectors and therefore identical except for the scaling factor k. The process of trying to find the maximum in the correlation function is equivalent to an experimental rotation of one of them until they become parallel.

The plots shown in Fig. 6-11 are merely the envelopes of the correlation functions. It should be pointed out that the sequence of two-dimensional plots is equivalent to a three-dimensional surface, continuous in both Δ and f_n. There is a



AMBIGUITY DIAGRAMS FOR SPECIFIC PULSE FORMS

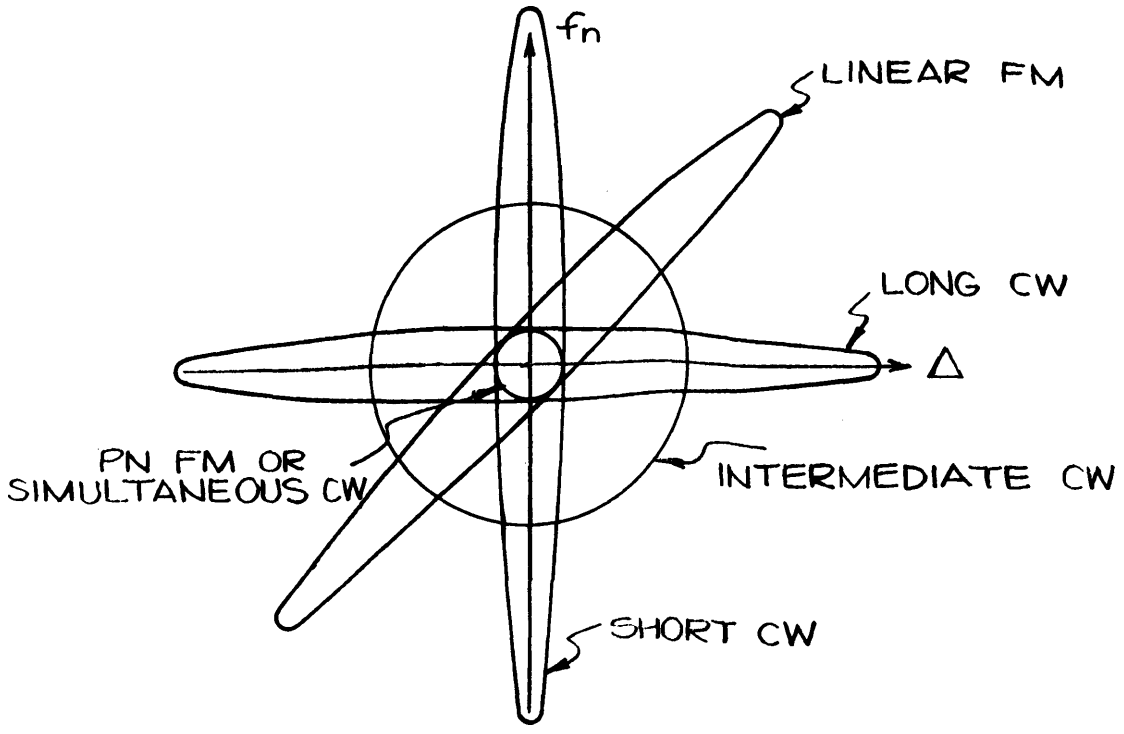


FIG. 6-11 - AMBIGUITY REGION

maximum of height k. The negative curvature in every direction from the peak depends upon the form of $u(f_o, t)$.

An important property of the surface is its ambiguity diagram. The ambiguity diagram is defined as the projection on the f, Δ-plane of the locus of points lying 3 dB below the peak. The simple circular region indicated in the upper part of Fig. 6-11 is somewhat special and is exhibited by specially constructed pulses. A number of ambiguity diagrams are shown in the lower part of Fig. 6-11. The short sinusoidal pulse possesses a cigar-shaped ambiguity diagram with its long axis parallel to the f-axis; the long sinusoidal pulse possesses a cigar-shaped ambiguity diagram with its long axis paralleled to the Δ-axis. A linear FM sweep possess a cigar-shaped ambiguity diagram with its long axis tilted with respect to the f and Δ axis. A sinusoidal pulse with length intermediate between the long and short pulse has an ambiguity diagram approximating a circle to the scale of the figure. If u(t) consists of a number of simultaneous sinusoidal pulses of different frequencies or a random frequency-sequence FM modulation, the ambiguity diagram may be made to be the small circle. Other pulse forms may not be so well behaved as these; a function $u(f_o, t)$ consisting of short bursts close together possesses a disconnected ambiguity diagram, for example.

From this brief description of the correlation function it can be imagined that a multi-channel correlator will have the largest output from the channel using the stored replica most like the signal being received. A correlator behaves like a filter which is sensitive to more than the power spectrum of the echo; it is sensitive to the details of the time behavior as well.

6.2 ENHANCEMENT OF SIGNAL-TO-NOISE RATIO

The example given in the introductory discussion showed that increasing the signal-to-noise ratio improved the ability

of the listener to detect the presence of a signal while it reduced his likelihood of making an error. This section will deal with methods of enhancing the signal-to-noise ratio. The conventional filter technique is well known as a way of discriminating against noise. The optimization of these filter techniques will be discussed in the following pages. The utilization of array gain is another familiar example of a technique for rejecting unwanted noise. The achievement of array gain was discussed as an example in a preceding section. The same principle applies in pre- and post-detection integration. Ping-to-ping integration is also a form of post-detection integration. As in the earlier parts of this discussion most of the concepts will be presented in terms of typical examples.

6.2.1 Conventional Filter Techniques. The principle of signal-to-noise ratio enhancement by filter techniques is well known. It will be briefly reviewed here as an introduction to the subject of optimum filtering. In this example a sinusoidal pulse of duration T is transmitted. The sonar receiver receives some of this energy scattered (1) from the sea surface, (2) from inhomogenieties in the bulk of the water, (3) from the ocean bottom, (4) from objects in the water such as marine life and target submarines. Except for reflections from distinguishable objects this received energy is called reverberation. In addition to this sonar induced acoustic energy there is ambient sea noise generated by wave motion, marine life, rain, etc., and noise caused by the motion of own ship through the water. Acoustic energy in these categories is called ambient noise or own ship's noise. It is the signal processor's job to find a way to enhance the acoustic signal reflected from a target submarine relative to the acoustic energy received from all other sources. An example of the kind of enhancement which can be accomplished is shown in Fig. 6-12. The upper channel shows the oscillograph trace of a waveform similiar to that which emerges

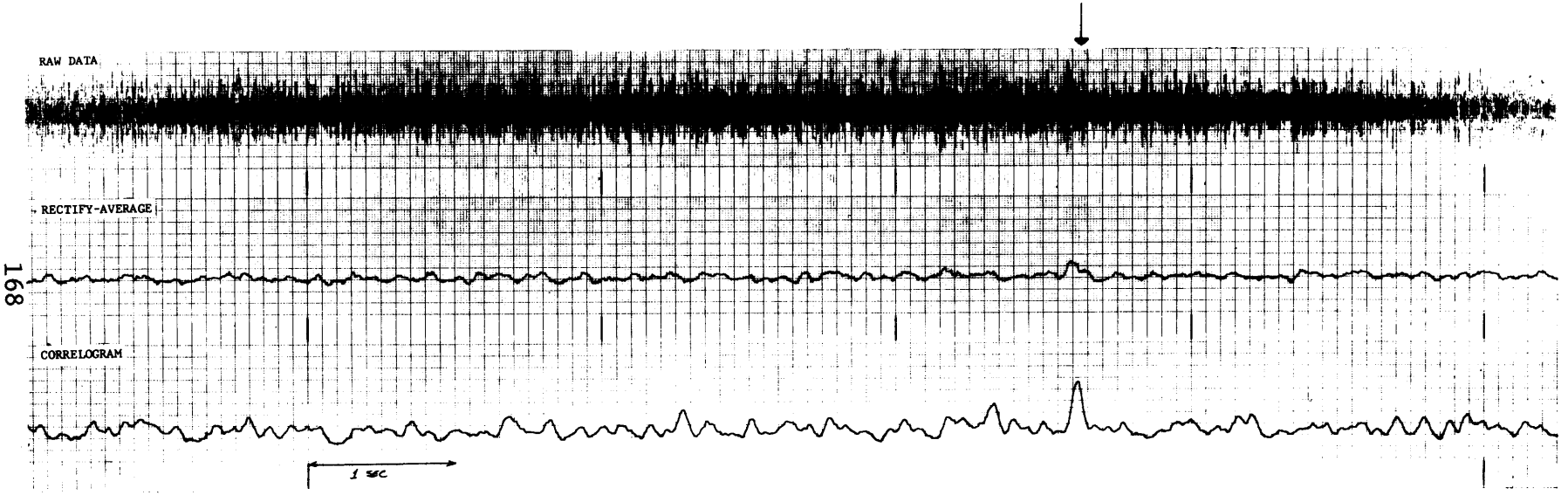
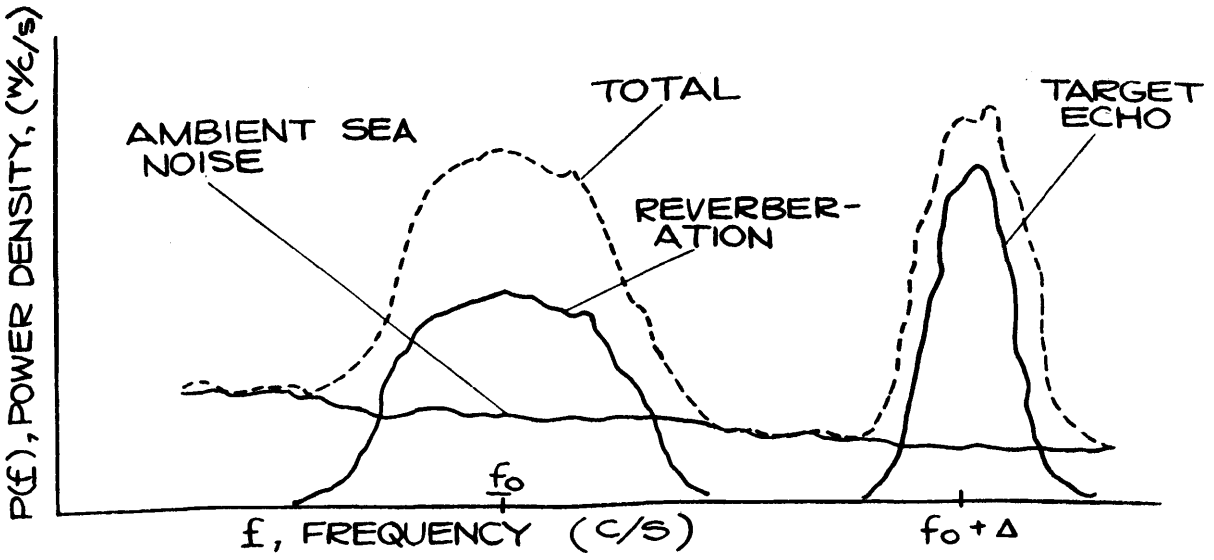


Fig. 6-12 Processor Output Signal in Noise (Pulse 70 ms, Noise BW = 200 c/s)

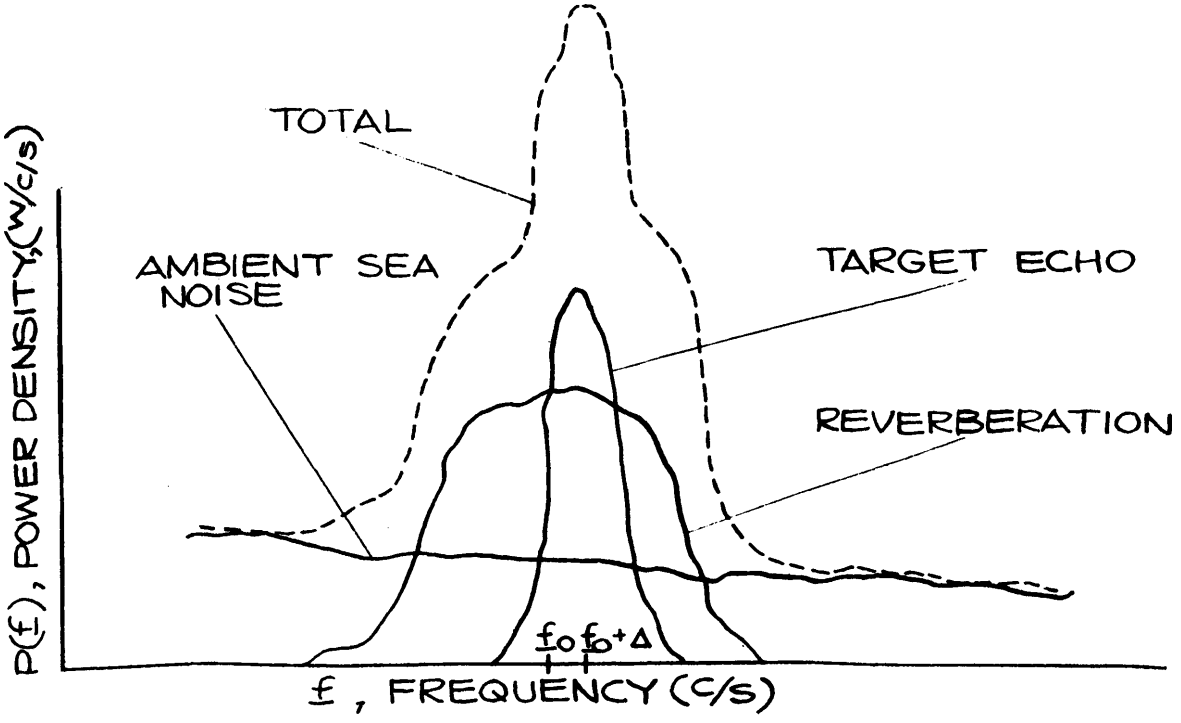
from a transducer. The arrow indicates the approximate position of the signal. This is a fairly large signal and it would probably be found without further filtering. The second channel shows the transducer output after rectification followed by filtering; the signal is perhaps more easily detected in this waveform. The third channel is obtained by filtering, then rectifying. The last process is one example of "optimum" filtering which will be discussed in a later section. The signal appears much more detectable in this channel. The signal-to-noise ratio seems to have been improved in this third channel.

The subject of improving the signal-to-noise ratio is, for the purpose of this example, most easily discussed in terms of the power spectra of the transducer output. Two typical spectra are shown in Fig. 6-13. The first represents the situation with an up-doppler target. The power spectrum at the output of the transducer is shown by the dotted curve. This spectrum may be thought of as being made up of three parts: an almost level plot arising from ambient sea noise, a component centered at f_o caused by reverberation, and a component centered at $f_o + \Delta$ caused by reflection from the up-doppler target. Δ is the doppler shift caused by the approaching target. The area under the "total" curve may be considered to be the total energy emerging from the transducer per unit time; it is then the average power from the transducer. When there is no signal present the component centered at $f_o + \Delta$ is not present. The background, or noise power, is the sum of the areas under the reverberation curve and the ambient noise curve. When the signal is present the area under the component centered at $f_o + \Delta$ is the signal power.

In order to make the signal more detectable a bandpass filter centered at $f_o + \Delta$ may be employed to pass the signal power and the noise which falls within its passband but reject



(a) A HIGH DOPPLER TARGET



(b) A LOW DOPPLER TARGET

FIG. 6-13 - SPECTRUM AT TRANSDUCER OUTPUT

TRACOR, INC. AUSTIN, TEXAS 1 DWG. 771-20 5-24-65 BMB-ES

all noise and reverberation outside its passband. In this way the interfering noise power is reduced without decreasing the signal power. As a result the ratio of the signal power to the noise power is increased.

A similar situation exists in the lower part of Fig. 6-13 except that the target echo falls within the reverberation band. The same techniques can again be applied but for this case the interfering "noise" will consist of power from the reverberation as well as from sea noise.

An important result can be obtained from this example. Suppose the echo is a sinusoidal pulse of duration T. Its spectrum is shown schematically in Fig. 6-14. The bandwidth $B_S = \frac{1}{T}$ of the signal contains nearly 90% of the signal energy ST. S is the average signal power while the signal is on. The area under the signal plot would equal to ST, the signal energy if the plot were made on a linear scale rather than a dB scale. The noise power density is η watts per c/s. The bandwidth B_n of the noise multiplied by η is the energy carried on the average by one second of the noise.

The signal-to-noise ratio $(S/N)_{in}$ at the input to the filter is

$$(S/N)_{in} = S/(\eta B_n) \qquad (6\text{-}63)$$

Suppose now that a bandpass filter centered on the signal is placed in the circuit so that a band B_s of the noise is passed. Almost 90% of the signal power is passed while only ηB_s of the noise power is passed. The signal-to-noise ratio $(S/N)_{out}$ at the output of the filter is

$$(S/N)_{out} = S/(\eta B_s) \qquad (6\text{-}64)$$

The 10% loss in signal power has been ignored in this example.

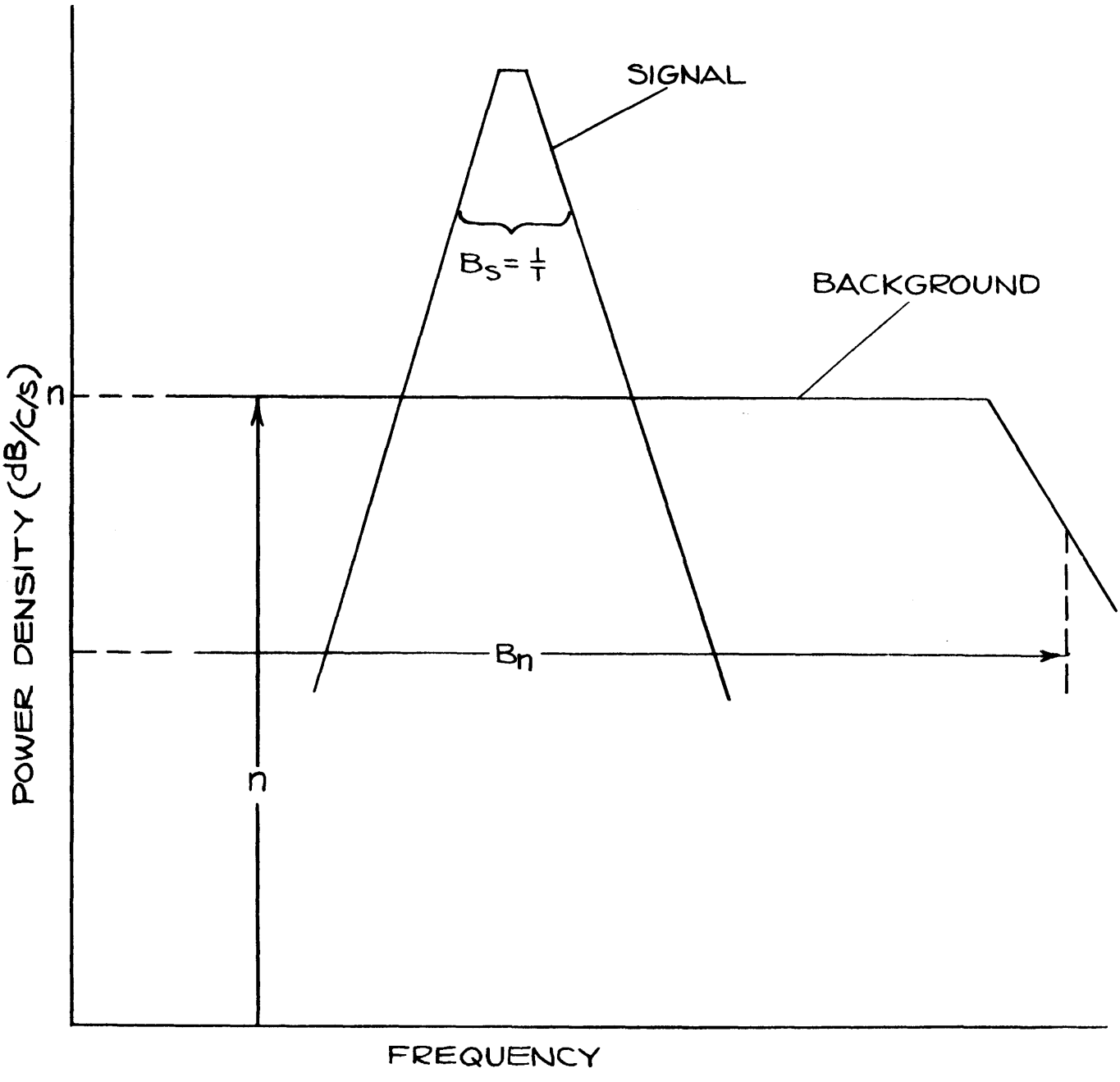


FIG. 6-19 - SPECTRA: CW SIGNAL AND UNIFORM BACKGROUND

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-21 | 5-24-65 BMB-ES

For a sinusoidal pulse it has already been stated that $B_s = 1/T$ so that Eq. (6-64) can be written

$$(S/N)_{out} = ST/\eta = E/\eta \quad . \tag{6-65}$$

In Eq. (6-65) S/N is average signal power output divided by the average noise power output. Often the peak signal output $\hat{S}$ is useful. For sinusoidal signals $\hat{S} = 2S$ so that

$$\left(\frac{S}{N}\right)_{out} = \frac{2E}{\eta} \quad . \tag{6-65a}$$

This expression states that the peak output signal-to-noise ratio from the filter is equal to twice the energy in the signal divided by the noise power per unit bandwidth. It turns out this is the largest possible output signal-to-noise ratio which can be obtained. As described here this filter is the optimum filter for this pulse form. In all signal processing systems the goal of the designer is to achieve this signal-to-noise ratio for the signal waveform being used.

Another concept often referred to in signal processing is processing gain. It is defined as the ratio of the signal-to-noise ratio at the output of the processor to the signal-to-noise ratio at the input of the processor. In this example, the processor was a bandpass filter. The processing gain is

$$G = \frac{(S/N)_{out}}{(S/N)_{in}} = \frac{ST/\eta}{S/B_n\eta} = B_nT \ , \tag{6-66}$$

the noise bandwidth at the input of the processor times the signal duration. This result holds for situations in which the noise background is "white," i.e., the spectral density of the noise is uniform. This expression, Eq. (6-66), is a source of much confusion in signal processing. It indicates that the larger the noise bandwidth the greater the

processing gain available. All this means is that if the input noise bandwidth is increased, there is just that much more noise power that can be eliminated by filtering. After all the filtering which can be accomplished is done, the output signal-to-noise ratio will be given by Eq. (6-65(a)). The job of the signal processor is still to obtain the output signal-to-noise ratio given by Eq. (6-65). A test that can be made to ensure success is to determine if the processing gain obtained agrees with Eq. (6-66).

The result in Eq. (6-65) can be derived in a more general way. This will be done in discussing the correlator as a signal processor. No new developments will be made except that a way will be described to provide optimum filtering for any pulse form. For pulse forms other than CW the processing gains and the optimum output signal-to-noise ratios will be in agreement with Eqs. (6-65) and (6-66).

This is a good place to digress shortly and discuss what happens in the example if the bandwidth of the filter is opened somewhat from the optimum. In the example as described a little more signal energy is passed, a maximum of about 10% more. The noise power on the other hand is doubled if the bandwidth is doubled. Since a 10% increase amounts to about a half a dB, this represents a degradation of the available output signal-to-noise ratio of nearly 3 dB. The filter is now processing the signal with a processing gain 3 dB below the optimum value.

The listener experiment described in the introduction can be explained with this model. It was found that the listeners detected sinusoidal signals of length T at 800 c/s embedded in noise of 210 c/s bandwidth as if they employed a bandpass filter with a 40 c/s bandwidth. A sinusoidal signal with a duration of 150 ms has a bandwidth of about 7 c/s, about one-sixth as wide as the 40 c/s bandpass. The listeners thus operate with 8 dB less processing gain than would be obtained with optimum

processing. An optimum processor in this case would have a passband of 7 c/s.

6.2.2 The Replica Correlator as an Optimum Filter

The properties of the correlation function provide a technique for implementing an optimum filter. It has already been shown that $\varphi_{uw}(f, \tau)$, the correlation function of u(t) and w(t), defined by

$$\varphi_{uw}(\tau) = \int_{-T/2}^{T/2} u(t)\, w(t + \tau)dt \ , \qquad (6\text{-}67)$$

is maximum where u(t) and w(t) are identical functions except for a scale factor and that the maximum occurs where $\delta = 0$. To illustrate how this filter action is obtained, the example of the sonar echo will be considered again; this time the background waveform will be considered as well.

As in the example discussed beginning on page 163, a function

$$u(f_o, t),$$

satisfying the condition

$$\sum_{i=1}^{n} u^2(f_o, t_i) = 1 \ , \qquad (6\text{-}68)$$

is transmitted. The function

$$ku(f, t - \frac{2r}{c}) \qquad (6\text{-}69)$$

is the scaled, doppler-shifted, and time-delayed echo from an ideal target or a single scattering center on a real target. This echo and the transmitted pulse have duration T. (Actually

the echo is stretched or compressed slightly by the target motion.) The echo frequency f is related to f_o by the usual doppler formula

$$f = f_o(1 - 2\frac{\dot{r}}{c}) \quad ; \tag{6-70}$$

$2\frac{r}{c}$ is the round trip travel time to the target, and k is determined by the target strength and two-way propagation losses.

In addition to the echo, there will be another function w(t) which is the background waveform - the noise or reverberation background in which the signal is embedded. Again a number of replicas of the transmitted pulse are produced with a number of doppler shifts.

$$\begin{aligned} &u(f_o, t) \\ &\quad u(f_1, t) \\ &\quad\quad u(f_2, t) \\ &\quad\quad\quad \cdots \\ &\quad\quad\quad\quad \cdots \\ &\quad\quad\quad\quad\quad \cdots \\ &\quad\quad\quad\quad\quad\quad u(f_n, t). \end{aligned} \tag{6-71}$$

Each of these is used as a "reference" function to "search" for the presence of the echo, $ku(f,t - 2\frac{r}{c})$, in the "noisy" waveform v(t):

$$v(t) = w(t) + ku(f,t - 2\frac{r}{c}) \ . \tag{6-72}$$

In general, w(t) is continually present while the echo has a shorter duration T and the time of occurrence of the echo is sought. The background waveform w(t) usually over-rides the echo so that it must be "sought" with as efficient means as possible.

The search will consist of laying each of the reference functions $u(f_n, t)$ alongside the function v(t) and multiplying the two functions point by point and summing the products. This is just the correlation function of $u(f_n, t)$ and v(t). The device used to perform this function will be referred to as the linear correlator and the output of the device will be called a linear correlogram. If this process is carried out over and over again as v(t) arrives, the complete correlation function $\varphi_{uv}(\Delta)$ for continuous Δ will be obtained.

$$\varphi_{uv}(\Delta) = \int_{\Delta}^{T+\Delta} u(f_n, t - \Delta)[w(t) + ku(f, t - 2\frac{r}{c})]\, dt, \qquad (6\text{-}73)$$

The range of integration Δ to $T + \Delta$ represents the time when the reference function is not identically zero. The echo is identically zero except in the interval $2\frac{r}{c} \leq t \leq 2\frac{r}{c} + T$. The background w(t) is a random function which is independent of u(t).

Outside the range where the signal exists, $\varphi_{uv}(\Delta)$ is simply

$$\varphi_{uw}(\Delta) = \int_{\Delta}^{T+\Delta} u(f_n, t - \Delta)\, w(t)\, dt \quad . \qquad (6\text{-}74)$$

When the correlation function is computed with sampled versions of u and w, Eq. (6-74) takes the form,

$$\varphi_{uw}(\Delta) = \sum_{i=1}^{i=j} u_{i-\Delta/\tau}\, w_i \tag{6-75}$$

In this sum τ is the time $1/(2B+\epsilon)$ between samples and j is the number of samples required in time T ($j = \frac{T}{\tau}$).

Since both u and w are functions with zero mean $\varphi(\Delta)$ also has zero mean. The square of φ is not zero, its average over Δ is

$$\overline{\varphi^2_{uw}(\Delta)} = \frac{1}{j} \overline{\sum_{i=1}^{j} u_{i-\Delta/\tau}\, w_i \sum_{k=1}^{j} u_{k-\Delta/\tau}\, w_k} \tag{6-76}$$

$$= \frac{1}{j} \overline{\left[\sum_{i=1}^{j} u^2_{i-\Delta/\tau}\, w_i^2 + \sum_{\substack{k=1 \\ i \neq k}}^{j} \sum_{i=1}^{j} u_{i-\Delta/\tau}\, u_{k-\Delta/\tau}\, w_i\, w_k \right]}$$

The second sum is a double sum which gives zero when averaged over Δ. The first is not zero and gives

$$\overline{\varphi_{uw}{}^2(\Delta)} = j\overline{u_i^2}\;\overline{w_i^2} \quad , \tag{6-77}$$

when averaged over Δ. According to Eq. (6-68), $\overline{u_i^2} = 1$ so that

$$\overline{\varphi_{uw}{}^2(\Delta)} = j\overline{w^2(t)} = 2BT\;\overline{w^2(t)} \; . \tag{6-78}$$

$\overline{w^2(t)}$ represents the average power in the background waveform; j is the number of statistically independent samples in the length T of the background. This number is of course $2BT + \epsilon$. This represents the average behavior of φ^2 when the echo is not present.

When the signal is present, its contribution is

$$\varphi_{uu}(\Delta) = k \int_{\Delta}^{T+\Delta} u(f_n, t-\Delta)\, u(f, t - \tfrac{u}{c})dt. \qquad (6\text{-}79)$$

The most interesting situation occurs for $f_n = f$ and $\Delta = \frac{2r}{c}$; then the two u functions are identical and

$$\varphi_{uu}(\Delta = \tfrac{2r}{c}) = k \int_{\Delta}^{T+\Delta} u^2(f, t - \tfrac{2r}{c})dt \qquad (6\text{-}80)$$

where the average is a time average over the signal interval of length T. Since $\overline{u^2(t)} = 1$, it follows that

$$\varphi_{uu}^2 (\Delta = \tfrac{2r}{c}) = k^2(2BT)^2 \qquad (6\text{-}81)$$

may be used as the signal power given by the correlation function. The noise power is given by Eq. (6-78) so that the continuous correlation function produces a power increase $\varphi_{uu}^2(\Delta = \frac{2r}{c})$ above the average noise power $\varphi_{uw}(\Delta)$ when the echo appears. It will be noted that

$$\frac{\varphi_{uu}^2(\Delta = \frac{2r}{c})}{\overline{\varphi_{uw}^2(\Delta)}} = 2BT \frac{k^2\overline{u^2(t)}}{\overline{w^2(t)}} \qquad (6\text{-}82)$$

Equation (6-82) may be rewritten by remembering that

$$k^2\overline{u^2(t)}$$

is the average signal power at the input to the correlator and that

$$Tk^2\overline{u^2(t)}$$

is the energy E of the signal, while

$$\overline{w^2(t)}/B$$

is the power per unit bandwidth η. With these combinations Eq. (6-82) becomes

$$\left(\frac{S}{N}\right)_{out} = \frac{\varphi^2_{uu}\left(\Delta = \frac{2r}{c}\right)}{\overline{\varphi^2_{uw}(\Delta)}} = 2\,\frac{E}{\eta} \quad . \tag{6-83}$$

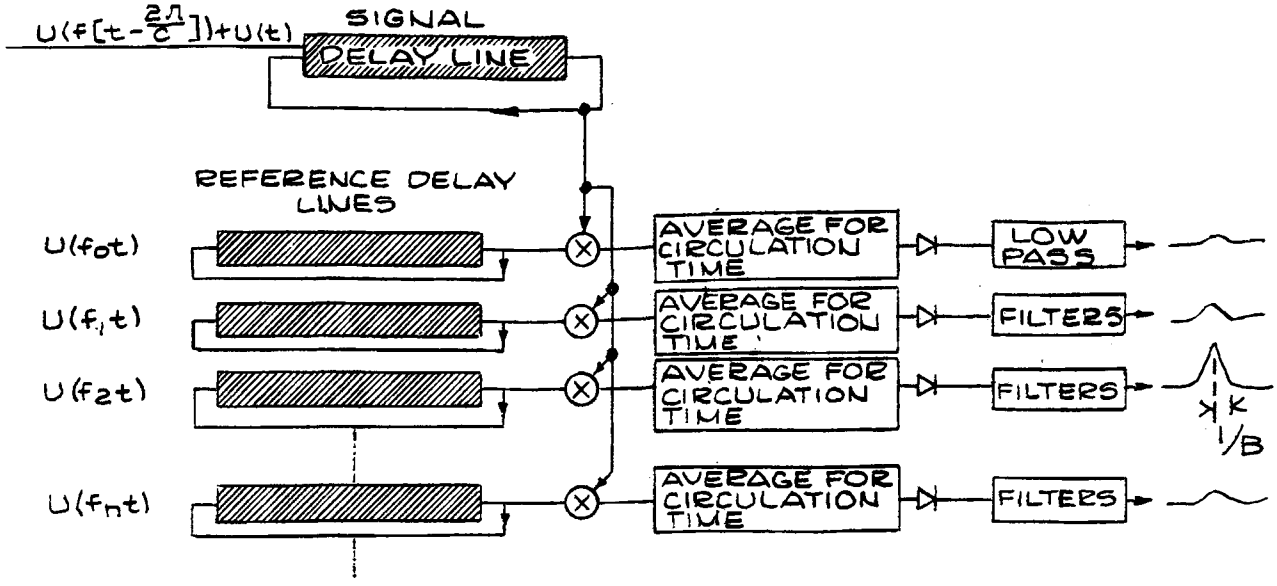
This result is identical to Eq. (6-65). The output signal-to-noise ratio is the same as that which was obtained for the CW pulse with conventional filter techniques.

6.2.3 Implementation of the Correlator

Practical implementation of the correlator is carried out by so-called time compression techniques. If a reference signal $u(f_n,t)$ has duration T, something more than 2BT independent samples must be multiplied by an equal number of samples of the incoming signal v(t) and summed before a new statistically independent sample is accepted from the receiver. This process can be accomplished in real time using digital techniques in either of two ways. The methods are known as the D.C. correlator and the heterodyne correlator. The D.C. correlator provides a direct analog of the correlation function and will be discussed first.

6.2.4 The D.C. Correlator

The block diagram of a D.C. correlator is shown in Fig. 6-15. The signal from the sonar receiver (background plus



TIME RESOLUTION = 1/B

FREQUENCY RESOLUTION = 1/T

LOW PASS FILTER CUT OFF = B
FREQUENCY SPACING = 1/T

FIG. 6-15 -INSTRUMENTATION: THE D.C. CORRELATOR

TRACOR, INC. AUSTIN, TEXAS
DWG. 771-46
6-25-65 BMB-ES

signal when it occurs) is inserted into the "signal delay line" in sampled form. It will be assumed that the delay lines are multi-bit devices which can store the time functions (signals, background, and reference functions) with high precision. The delay lines are "time compression" devices. For example, the signal line contains a number of sequential samples of the sonar receiver output, taken at a rate in excess of $2B + \epsilon$ ($\epsilon>0$), twice the maximum frequency, f_{max}, in the data channel. The samples in the delay line represent the output of the receiver over the latest T seconds; the time τ between samples is somewhat less than $\frac{1}{2f_{max}}$ seconds. The contents of the delay line circulate around the loop continually and emerge from the delay line each circulation. The time required for one circulation is τ, identical to the sampling period. Samples representing T seconds of receiver output are, in this way, made available for display or computation each τ seconds. The delay line therefore provides a time compression factor $\frac{T}{\tau}$.

The waveform in the signal delay line is updated at each circulation by deleting the oldest sample and inserting a new sample from the receiver.

The reference delay lines are loaded with the series of doppler-shifted replicas of the transmitted pulses of Eq. (6-71). These references also circulate around the delay line each τ seconds. The samples from each of the reference functions are fed to a sequence of multipliers (correlators) where they are multiplied by the samples from the signal delay line. The products are then averaged for the circulation time τ, which corresponds to averaging for the duration of the signal, T, in a non time-compressed system. The outputs from these averagers are linearly rectified and introduced into a low pass filter whose bandwidth is B. This filter may be implemented by averaging for $\tau_1 = \frac{1}{B}$. As time passes, the outputs of the filters represent their response to background until the echo $u(f, t - \frac{2r}{c})$ enters the delay line and matches up with the reference function most like it. At this time a signal spike appears at the output of one of the filters. Smaller spikes may appear at the

output of the other filters if the pulse form has poor doppler discrimination capabilities.

The preceding description has given the behavior of the circuit. It has many features similar to the characteristic of the correlation function. One additional consideration needs to be brought out to show, for example, that the signal peak comes from a single filter output and has a duration of $\frac{1}{B}$ seconds.

The upper part of Fig. 6-16 shows a schematic representation of the transmitted waveform, in this case a form of pseudo-noise FM. The solid plot in the second part of the figure contains the pulse frequency plotted as a function of time; this plot can also be considered to be the frequency of one of the reference functions $u(f_n,t)$ as a function of time. The dotted plot represents the frequency of the ideal echo plotted as a function of time; the two functions are in perfect time registry for this example.

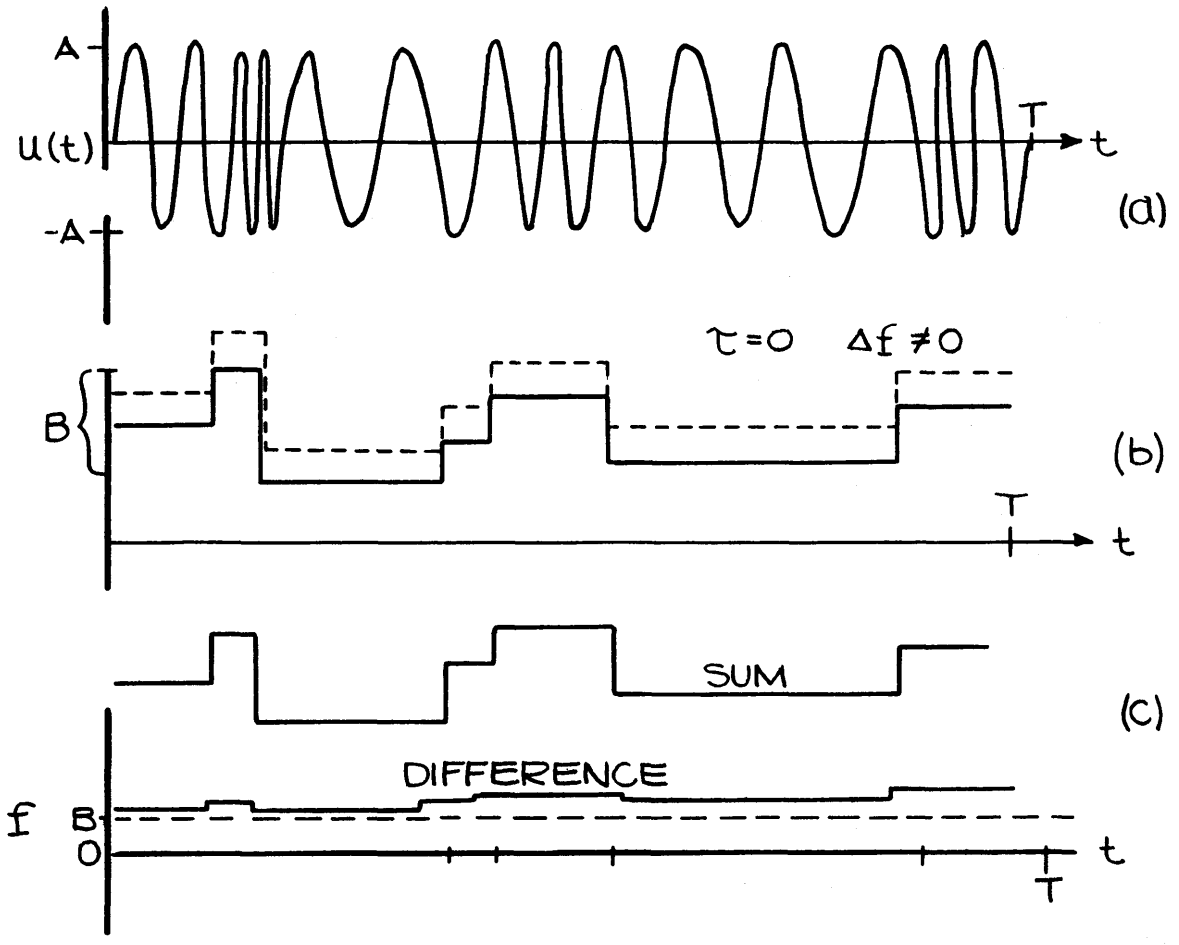
The correlator is a multiplying device which multiplies the two time functions whose frequencies are represented in Fig. 6-16(b). During the time the frequency is constant, two time functions of the form

$$\sin(2\pi ft) \times \sin(2\pi f_n t)$$

$$= \frac{1}{2}[\cos 2\pi(f-f_n)t - \cos 2\pi(f+f_n)t] \qquad (6\text{-}84)$$

are multiplied with the result shown. In Eq. (6-84), f is the frequency in the echo; f_n is the frequency in the reference. At the output of the multiplier one cosinusoidal wave appears at the difference frequency and another at the sum frequency. As the reference frequency jumps discontinuously from one value to another, the signal frequency also jumps in the same direction by nearly the same amount.[4] Because of the time compression these frequencies are the time-compressed values.

[4]The differences which might arise are less than $2B\dot{r}/c$, where B is the pulse bandwidth, $\dot{r}$ is the target range rate and c is the speed of sound.



OUTPUT IS CW:

FREQUENCY RESOLUTION $\sim \frac{1}{T}$

FIG. 6-16- FREQUENCY RESOLUTION: PN FM (D.C. CORRELATOR)

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-47 | 6·28·65 BMB-E

The difference frequency $f-f_n$ is therefore nearly constant throughout the period T. The sum frequency on the other hand jumps from one frequency to another with twice the size of the discontinuity gap of the component frequencies. It is apparent that a continuous single frequency sinusoidal pulse results from the difference frequency term while a variable frequency pulse results from the sum term (Fig. 6-16(c)).

The output of the correlator is passed through a low-pass filter with an averaging time corresponding to the sampling interval, τ. All the frequencies are being discussed as if they were at their normal values, but it must be remembered that they are really at the compressed value $\frac{T}{\tau}(f_n-f)$. For this to be less than $\frac{1}{\tau}$, the inequality,

$$\frac{T}{\tau}(f_n-f) \leq \frac{1}{\tau} , \qquad (6\text{-}85)$$

must be satisfied. This is equivalent to

$$T(f_n-f) \leq 1 , \qquad (6\text{-}86)$$

It thus follows that

$$f_n-f \leq 1/T , \qquad (6\text{-}87)$$

if the correlator is to get a signal passed through the low-pass filter to the output detectors. All of the reference signals whose frequency separation from the echo is greater than that given by Eq. (6-87) provide little or no output from their low-pass filters. Of course all of the energy at the sum frequency will fall outside the low-pass filter pass-band. This fact leads to the conclusion that the D.C. correlator provides frequency resolution determined by the reciprocal of the pulse length.

The time resolution provided by this D.C. correlator will be considered for the channel with the reference and echo doppler matched. Figure 6-17(a) has as its transmitted pulse a pseudo-noise frequency modulation as in the previous example. In Fig. 6-17(b) the instantaneous frequency plots are shown again for the case where the echo and reference are time-matched. It is clear that the difference between the reference frequency (solid line) and the signal frequency (dotted line) is zero and the difference frequency term out of the correlator will be at zero frequency and be passed by the low-pass filter. The time resolution is determined by the amount one of these functions can be displaced before the difference frequency falls outside the passband of the low-pass filters a significant fraction of the time.

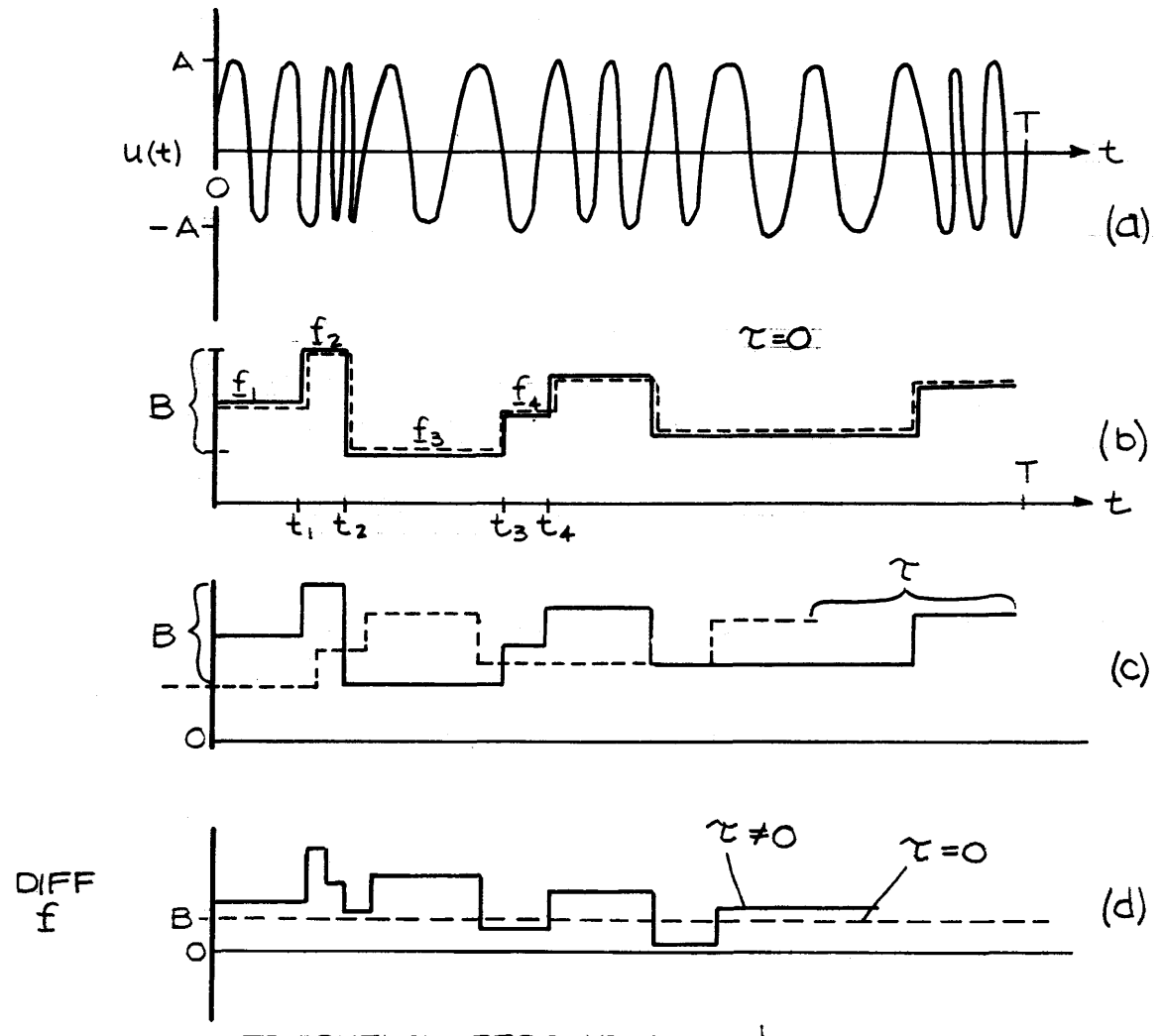
The frequency resolution of the system is $\frac{1}{T}$; in a bandwidth B, $k = \frac{B}{1/T} = BT$ resolvable, discrete frequencies can be picked from the pulse. The average time per frequency is $\frac{T}{k} = \frac{1}{B}$. The time shift which can occur before significant mismatch of the echo and reference frequencies is therefore about $\frac{1}{B}$. It follows then that two arrivals can be recognized as discrete if their time separation $t_2 - t_1$ satisfies the inequality

$$t_2 - t_1 \geq \frac{1}{B} \quad . \tag{6-88}$$

This is a crude justification of the time resolution behavior; a similar and more convincing demonstration can be made with the Fourier series representation of the time functions.

6.2.5 The Heterodyne Correlator

In the previously described correlator the difference frequency term in the correlator product was the source of the output signals, the sum frequency terms being blocked by the filters. In order to implement this correlator a number of correlator references were required. There is a second type of



FREQUENCY RESOLUTION = $\frac{1}{T}$
CAN PICK k FREQENCIES: $k = B/(1/T)$
AVERAGE TIME PER FREQUENCY = $\frac{T}{k} = \frac{1}{B}$
τ TO MISMATCH $\sim \frac{T}{k} = \frac{1}{B}$

TIME RESOLUTION $\sim \frac{1}{B}$

FIG. 6-17 - TIME RESOLUTION : PN FM (D.C. CORRELATOR)

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-48
6-25-65 BMB-ES

correlator, known as a heterodyne correlator, which makes use of the sum frequency component and requires only one reference channel. Its block diagram is shown in Fig. 6-18.

The signal delay line operates as before; the oldest sample is deleted and a new sample added at each circulation. The contents of the delay line are fed to the multiplier one sample at a time during the circulation. In this version of the correlator the reference delay line contains a spectrum inverted replica of the transmitted pulse $u(f_o,t)$. A spectrum inverted replica is constructed by heterodyning $u(f_o,t)$ against a reference oscillator above the spectral band occupied by $u(f_o,t)$.

An example will illustrate this procedure. Suppose $u(f_o,t)$ is made up of three frequencies: a center frequency f_o, a frequency at $f_o - \Delta$, and a frequency $f_o + \delta$. Suppose $u(f_o,t)$ is multiplied by a continuous sinusoid at frequency $2f_o$. The frequencies present in the product will be $2f_o$ plus each of the frequencies in $u(f_o,t)$ and $2f_o$ minus each of the frequencies in $u(f_o,t)$:

$$
\begin{aligned}
|2f_o + (f_o - \Delta)| &= 3f_o - \Delta \\
|2f_o + f_o| &= 3f_o \\
|2f_o + (f_o + \delta)| &= 3f_o + \delta \\
|2f_o - (f_o - \Delta)| &= f_o + \Delta \qquad (6\text{-}89) \\
|2f_o - f_o| &= f_o \\
|2f_o - (f_o + \delta)| &= f_o - \delta \ .
\end{aligned}
$$

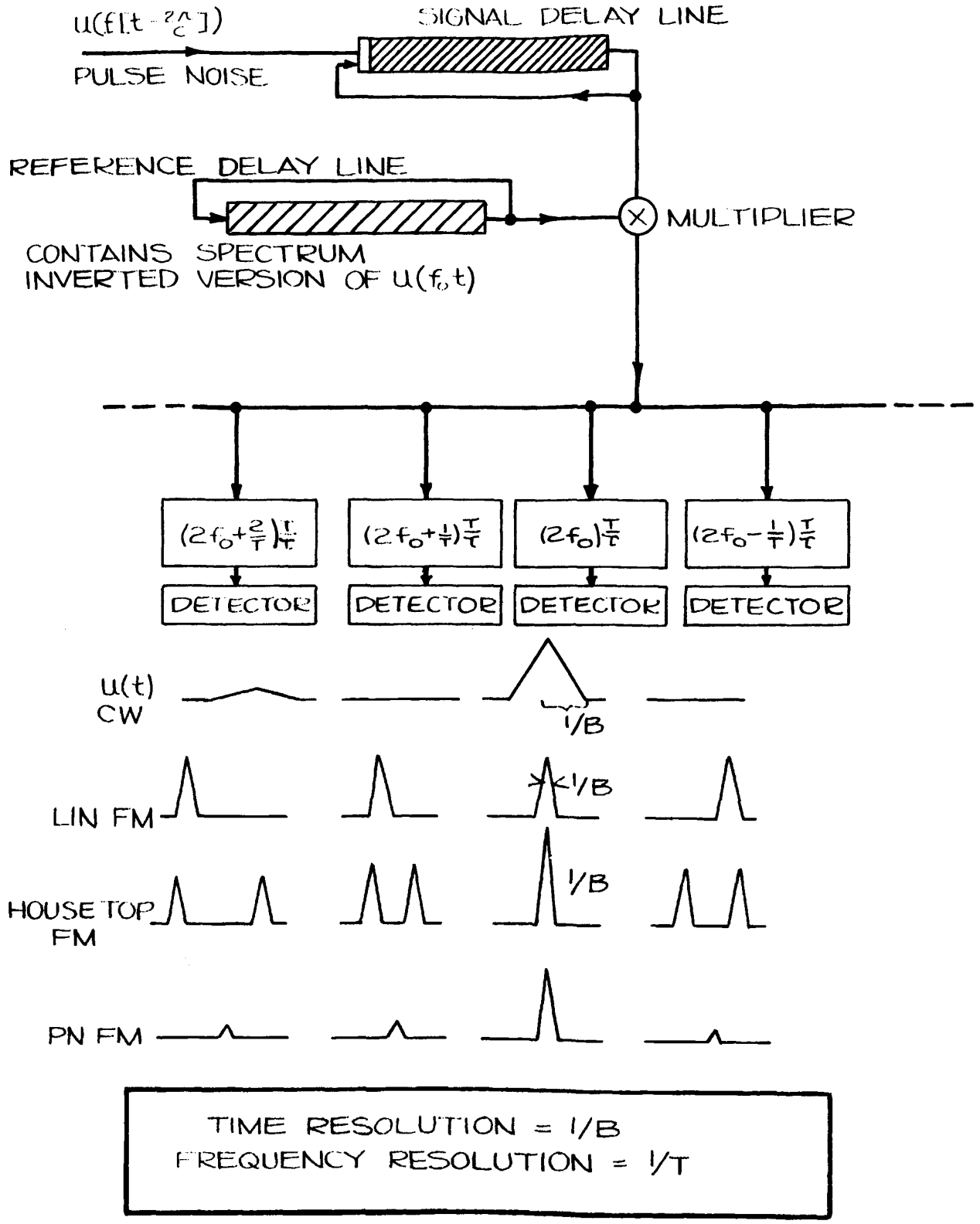


FIG. 6-18- INSTRUMENTATION: THE HETERODYNE CORRELATOR

TRACOR, INC. AUSTIN, TEXAS
1 DWG. 771-49
6-28-65 BMB-ES

The upper three frequencies are separated from the lower three by filtering. The transformation which has occurred is

Original Spectral Components	Transformed Spectral Components
$f_o - \Delta$	$f_o + \Delta$
f_o	f_o
$f_o + \delta$	$f_o - \delta$.

(6-90)

The transformed spectrum is said to be spectrum inverted. Those spectral components which were originally Δ below f_o are now Δ above f_o, etc. It is necessary that this type of spectral inversion be applied to the reference signal before the reference delay line is loaded.

Except for doppler shift $2 f_o \dot{r}/c$ on each of the original spectral components the signal delay line contains just the original spectral components while the reference delay line contains the transformed components. When the outputs of the two delay lines are multiplied sample by sample the resulting function contains the sum and difference of all frequency pairs in the two stored functions.

Signal Spectral Components	Reference Spectral Components	Product Spectral Components
$(f_o-\Delta)$	$f_o+\Delta$	$2f_o$
f_o	f_o	$2f_o$
$(f_o+\delta)$	$f_o-\delta$	$2f_o$
		$\Delta, 2\Delta, 2f_o \pm \Delta$
		$\Delta \pm \delta, 2f_o \pm (\Delta+\delta)$
		$\delta, 2\delta, 2f_o \pm \delta$

It will be noted that all of the corresponding terms are translated to $2f_o$ while the cross terms are separated from $2f_o$ by about $2f_o$ Δ,δ, or $\Delta+\delta$. The terms not containing $2f_o$ are separated from those containing $2f_o$ components by means of filters; the remaining terms near frequency $2f_o$ are sinusoidal pulses and can be filtered in an optimum fashion by means of the narrow band filters centered at $(2f_o)\frac{T}{\tau}, (2f_o \pm \frac{1}{T})\frac{T}{\tau}, (2f_o \pm \frac{2}{T})\frac{T}{\tau}, (2f_o \pm \frac{3}{T})\frac{T}{\tau}$, ... (Fig. 6-18).

It should be remembered that the frequencies of the signals appearing at the output of the delay lines are scaled up from those of the real time signals by the time compression ratio. Typical frequencies at this point might be as high as 500 kc while the corresponding real time frequencies are only about 300 c/s. Also the doppler shift has been omitted from the discussions; the three frequencies near f_o do in fact differ slightly from each other, their actual values are altered by the factor

$$(1 - \frac{2\dot{r}}{c}) \quad .$$

The frequencies separated from $2f_o$ by $\pm\Delta$ and $\pm\delta$ fall in one of the adjacent filters, one spectral component contributing to each one. The frequencies $2f_o$ are contributed by each spectral component and all appear in the filter labeled $2f_o$. If there are n spectral components present, the output of the $2f_o$ filter will be n times as large as the average output from the other filters.

The typical outputs shown below the filters represent detected versions of the outputs from the filters for a few common pulse forms. Comparison of these outputs with the ambiguity diagrams of Fig. 6-11 will show that the behavior is exactly as expected. The CW pulse has good doppler resolution but poor range resolution; the FM pulse has good range resolution but the doppler shift appears as a range shift; the housetop FM exhibits a double peak separated by an amount proportional to doppler shift; the pseudo noise FM exhibits simultaneous time and frequency resolution.

Figure 6-19 shows the signal spectrum f, the inverted reference spectrum $\check{f}$, and the product spectra $\check{f}+f$ and $\check{f}-f$ for the pseudo-noise FM pulse. The manner in which the CW signal is formed at the sum frequency output of the multiplier is easily visualized in this diagram.

6.3 INFORMATION EXTRACTION

In general, one can hope to accomplish one or more of the following by means of signal processing:

Detection - establishing the presence or absence of an echo.

Localization - determining the range, azimuth and range rate of the object producing the echo.

Classification - recognition of echo structure characteristic of a submarine target.

In this section a discussion of how these three tasks might be performed with a correlator will be presented.

6.3.1 Detection

In order to optimize the detection procedure it is usually considered desirable to increase the signal-to-noise ratio. The processor which provides the maximum output signal-to-noise ratio is a form of the replica correlator. It provides a peak signal S to average noise N ratio amounting to

$$\frac{S}{N} = \frac{2E}{\eta} \tag{6-91}$$

in which E is the energy in the signal to be detected and η is the noise power per unit bandwidth.

Let us consider how a sonar system might be modified to increase the signal-to-noise ratio. Obviously an increase in

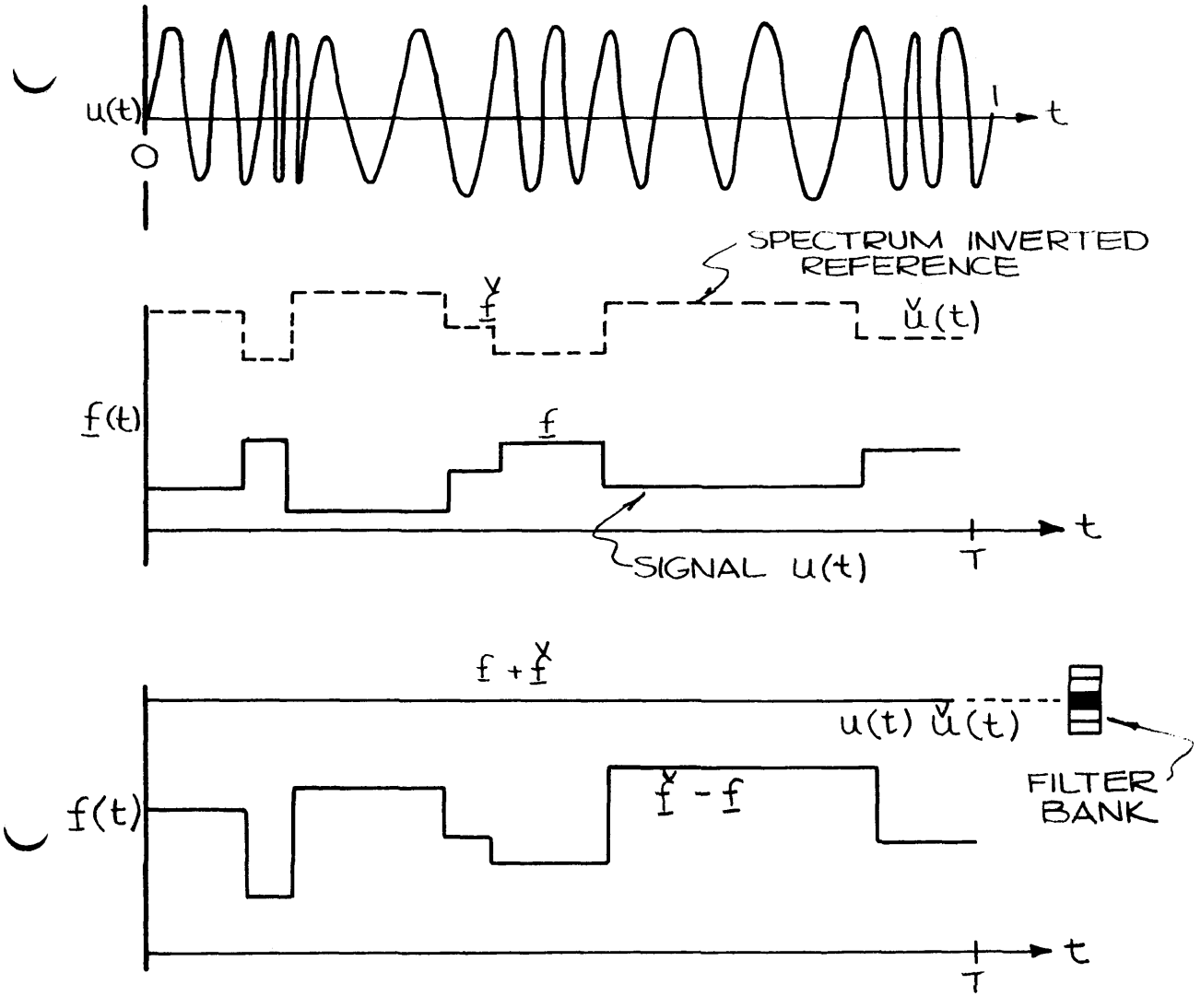


FIG. 6-13 - THE HETERODYNE CORRELATOR

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-50 | 6-28-65 BMB-ES

signal energy is a step in the right direction. For an active sonar there are three measures which can be applied to increase E:

1. Raise the level of the transmitted pulse
2. Increase the directivity index of the transmitting array
3. Increase the length of the transmitted pulse.

The first and third of these measures get more energy into the water, hence more energy strikes the target and more energy is carried by the target echo. Increasing the directivity index of the transmitting array does not put more energy in the water but does localize more energy in the direction of the target; therefore more energy is reflected from the target.

As these measures are applied, limits are reached beyond which no further gains can be obtained by these measures alone. As the level of transmission is increased, cavitation at the transducer face sets in and limits the power which can be transformed into acoustic waves. An increase in directivity index involves enlarging the transducer aperture or transducer diameter. Clearly this will be a limiting factor when it is remembered that the transducer enclosure alters the flow pattern around the ship thereby degrading ship performance. It should also be pointed out that narrowing the sonar beam increases the number of search beams which must be programmed to search a given volume of the ocean. Finally, increasing the pulse length seems the most attractive means for increasing E. This expedient also has its limitations. When the pulse length is made sufficiently long the reverberation level increases until at a particular value of pulse length the reverberation background becomes equal to the ambient noise background and from this point on the quantity η begins to depend upon reverberation rather than ambient noise. Increases in pulse length beyond this limit increase E and η proportionally so that no further change in signal-to-noise ratio can be achieved.

The other approach to improving the available signal-to-noise ratio is the reduction of η. This may be accomplished in two ways:

1. Increase the receiving directivity index.
2. Increase the bandwidth of the transmitted pulse.

The first measure is limited in the same way as the attempts to increase the transmitting directivity index. In addition this measure tends to raise the reverberation level relative to the ambient noise level and to promote reverberation limiting. The second measure is designed to reduce η when the background waveform consists of reverberation rather than ambient noise. If, for example, the pulse length has been increased until the reverberation level per unit bandwidth is just equal to the noise level per unit bandwidth the reverberation per unit bandwidth can be halved by doubling the bandwidth of the transmitted pulse. At first glance this technique offers the cure to all sonar ills because it seems that the pulse length could be increased indefinitely if the bandwidth is raised proportionally. This results in an increase in E while η remains constant. However, this measure is also restricted in practice by (1) the available passbands of transducers, (2) a phenomenon known as multipath energy splitting, and (3) the inherent output bandwidth of the optimum processor.

Suppose now that all of the available measures for increasing E relative to η have been applied. It is necessary to provide an optimum processor to achieve the maximum possible output signal-to-noise ratio (See Eq. (6-91)) for use in detection. For high doppler targets, the reverberation and echo are separable in frequency by means of filters. An optimum system can be designed which utilizes long CW pulses whose echoes are passed through a set of comb filters with bandwidths equal to the reciprocal of the pulse length. (See Fig. 6-13) This type of system is the best attainable from the standpoint of both the maximum signal-to-noise ratio and the processor output bandwidth.

The problem areas lie in the detection of the low doppler target where the target echo and reverberation fall in the same frequency range (See Fig. 6-13). It is for this case that the wide band pulses are useful. Optimum processing is most easily carried out on these pulses with the replica correlator. When an FM-sweep transmission is used the output signal-to-noise ratio obtained is almost independent of target doppler.

Multipath. The measures employed to increase the available signal-to-noise ratio (Eq. (6-91)) resulted in an increase in the signal bandwidth. Since range resolution δr depends upon the time resolution δt of the pulse and δt is determined by the bandwidth, it follows that target structure will begin to be resolved as bandwidth is increased. Two discrete arrivals can be distinguished if their time separation t_2-t_1 satisfies the inequality

$$t_2 - t_1 \geq \frac{1}{B} \tag{6-92}$$

(See Eq. (6-88)). It follows that two objects separated in range by $r_2 - r_1$ can be distinguished as discrete objects provided the round trip time from r_2 to r_1, exceeds $\frac{1}{B}$, hence they can be resolved if

$$r_2 - r_1 \geq \frac{c}{2B} \quad . \tag{6-93}$$

For a bandwidth of 100 c/s, $r_2 - r_1$ satisfies this inequality if it exceeds 25 ft. For this bandwidth, discrete scattering centers on the target are resolvable. The target therefore produces a number of "echoes" close together rather than a single echo. At least this is the appearance at the output of the correlator. An example of this behavior is shown in Fig. 6-20.

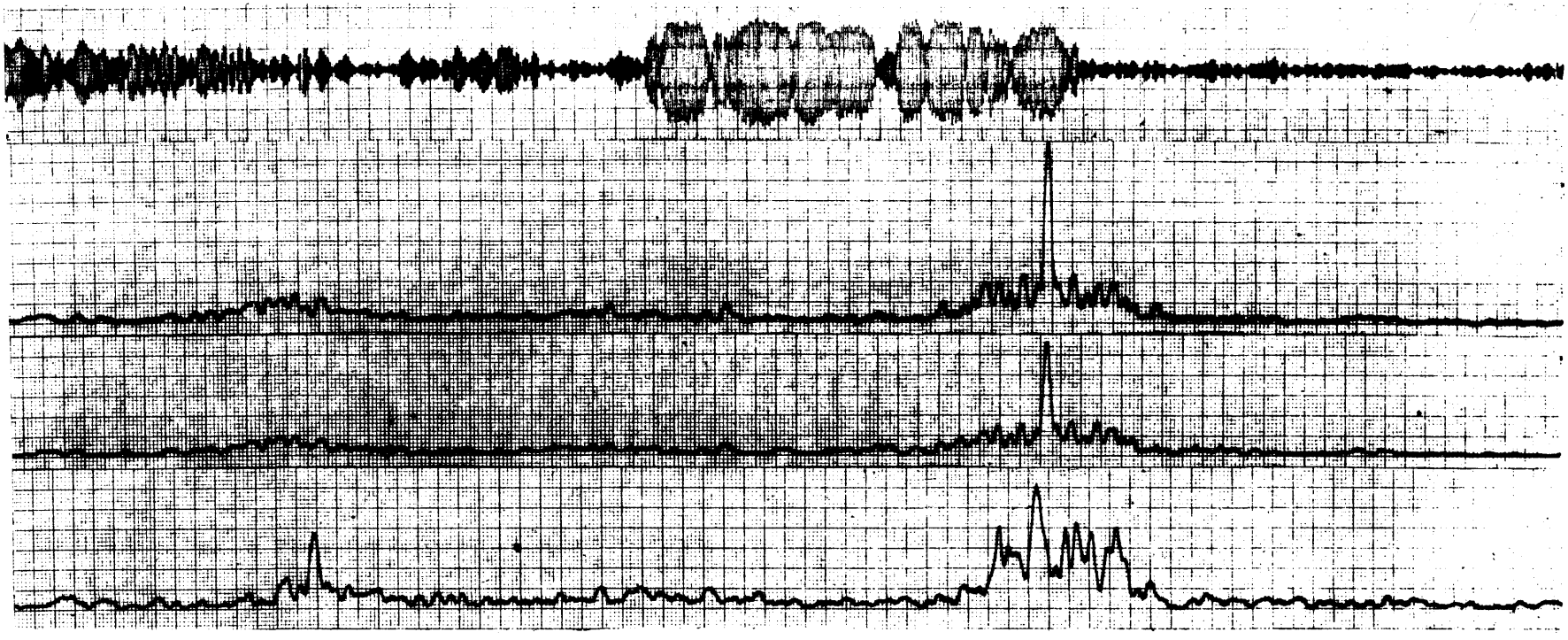


Fig. 6-20 An Optimum and a Non Optimum Processor

The upper channel shows the echo obtained from a submarine target using a 200 ms, 100 c/s FM sweep. The lower channel shows the output of a D.C. correlator obtained with this echo using a replica of the transmitted pulse as a reference signal. Clearly the energy in the correlator output is distributed between several resolved output spikes. The target consists in effect of several discrete scatterers. If the transmitted pulse reference signal is replaced with a signal similar to the actual echo, the result in the second channel is obtained. Here practically all of the energy of the echo is stacked neatly in a single spike. The third channel is identical to the second except that the vertical gain is lower. The second and fourth channel outputs were both obtained with replica correlators; only the second channel output can be considered an approach to optimum processing because a reference signal like the echo was employed to produce this correlogram.

6.3.2 Localization

The aim in localization of the target is to determine its position in space (its range, azimuth, and depth) and its range rate. Both range and range rate can be determined by means of a correlator output provided the range-range rate ambiguity of the transmitted pulse is suitable. In the past these two functions have been performed separately. Range rate has been determined by doppler methods by using CW transmissions and suitable narrowband filters for determining the doppler shift of the echo. Range measurements have been based upon the time of arrival of either a short pulse or the time of occurrence of the output peak of a real time correlator. The transmission employed with the correlator is usually an FM slide. The simultaneous measurement of range and range rate using pseudo-noise pulses and a linear correlator looks attractive; demonstration of a practical system for accomplishing this is as yet an unattained goal.

6.3.3 Classification

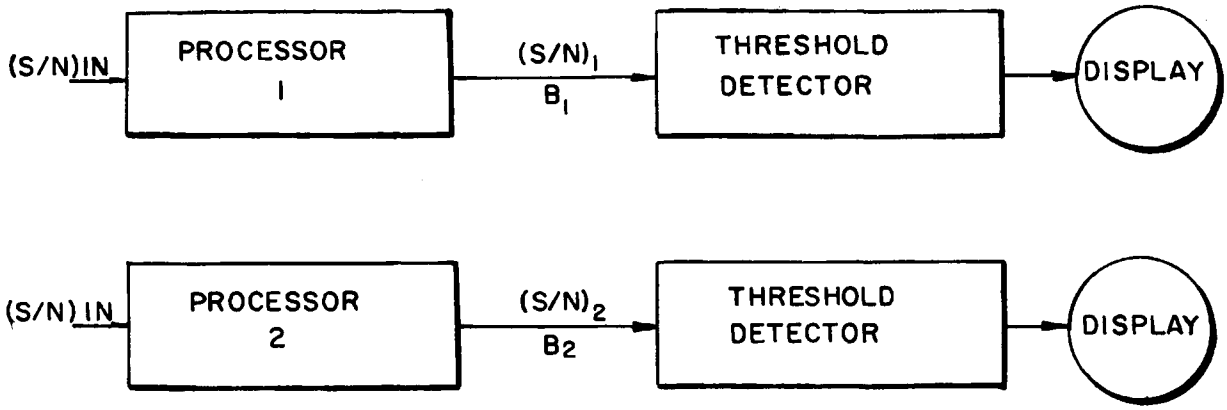
The lower channel of Fig. 6-20 indicates correlogram structure related to the echo formation process. Each spike in the correlogram is presumably related to some structural feature of the target. With sufficiently high resolution the correlogram details can be identified with hull structural details and be used as a basis for determining whether or not the target is in reality a submarine. The correlator with a transmitted pulse reference provides an output spike for each resolvable scattering center on the target. The structure observed in the correlogram is very similar to the structure usually found in short pulse echoes.

6.4 COMPARISON OF DETECTION PERFORMANCES OF PROCESSING SYSTEMS

6.4.1 Signal-to-Noise Ratio and Output Bandwidth

In the preceding sections methods of increasing the output signal-to-noise ratio to the optimum value $2E/\eta$ were discussed. The simple example concerning the listeners described in the introduction indicated the advantage of higher signal-to-noise ratios. This advantage is immediately apparent for any specific processor; detection is easier when the signal-to-noise ratio is increased.

This situation is not so straightforward when two different processing systems are to be compared. To illustrate the proper method of comparison of the detection performances of two systems, the discussion will be restricted to the class of so-called "threshold detectors." The threshold detector is a decision element; it provides an output when the waveform being examined exceeds the threshold and provides zero output otherwise. This behavior is known as its "decision rule." The systems to be compared are represented in block form:



The systems are identical except for the processors employed. They are assumed to have identical signal-to-noise ratios at their inputs. If the two processors are not equivalent, the signal-to-noise ratios at the outputs may be different. The processor output is then introduced to the threshold detector whose output is either 0 or +1. The display indicates the presence of +1 outputs from the threshold detector.

It may happen that two processors which have the same output signal-to-noise ratios do not have the same detection performance when used with threshold detectors. The difference in their performances arises because of the difference in the output bandwidths of the processors.

At the end of Section 6.1.1 a theorem called the "Time Resolution Theorem" appears. It states, in effect, that every $\frac{1}{B}$ seconds, on the average, a sample of noise statistically independent of the preceding sample arrives. B is the bandwidth of the noise. The average number of independent noise samples which arrive per second is therefore B. When noise alone is present in the output of the processor there are therefore B independent opportunities per second to actuate the display with the

threshold detector on noise alone. If G(0,x) is the probability that a noise waveform alone exceeds the threshold x in a particular sample of noise, then the average number F of "clutter" points on the display per second is

$$F = BG(0,x) \quad .$$

The threshold x is adjustable in the threshold detector so that F can be controlled and set at any desired rate. When a signal arrives, the signal plus noise will have a higher probability $G(\frac{S}{N},x)$ of exceeding the fixed threshold x than will noise alone. Here $G(\frac{S}{N},x)$ is the probability that the signal plus noise will exceed x. $G(\frac{S}{N},x)$ is called the detection probability. It should be pointed out that when the signal lasts longer than the average time between independent samples of noise that the signal will be embedded in more than one sample of noise and there will be more than one opportunity to detect the signal.

A simple example will show why processors cannot be compared in terms of their output signal-to-noise ratios alone. Suppose $(\frac{S}{N})_1$ and $(\frac{S}{N})_2$, the output signal-to-noise ratios of the two processors, are identical, that the waveform statistics of both processors are described by $G(\frac{S}{N},x)$ and that the same threshold setting has been made in both threshold detectors. When this has been done, the probability of detecting the signal, when it is present, will be the same in the two processors. Now if the output bandwidths of the two processors are different, such that

$$\frac{B_1}{B_2} = k \qquad (6\text{-}94)$$

then the expected clutter rates will be

$$F_1 = B_1\ G(0,x)$$

$$F_2 = B_2\ G(0,x) \qquad (6\text{-}95)$$

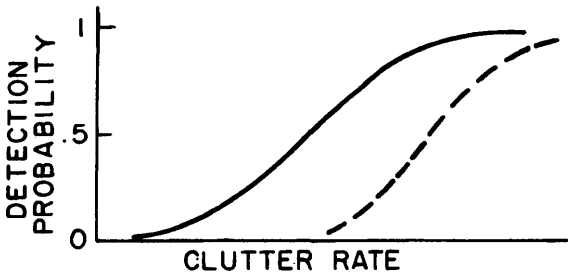
and $F_1 = F_2\ (B_1/B_2) = k\ F_2$, so that the clutter rate on the display of processor No. 1 will exceed that on the display of processor No. 2 by a factor k. There will be k times as many spurious points on the display of the processor with the larger output bandwidth. Clearly, the No. 2 processor is the better of the two in this comparison, yet the processors provided the same output signal-to-noise ratio.

The difference in clutter rates could be removed by raising the threshold x_1 of the threshold detector of the No. 1 processor. The processors will obviously not be equal in the presence of a signal because the higher threshold x_1 necessarily decreases the probability that signal plus noise will exceed the new threshold. It follows then that the detection probability obtained with the No. 1 processor falls below the detection probability obtained with the No. 2 processor.

As a result of this example it may be assumed that of two processors with different output bandwidths, the one with the greater bandwidth must have greater processing gain in order to compete with the performance of the narrow band processor.

In the preceding example the statistical properties of the output waveforms were assumed to be identical. In general, it is necessary to compare two processors in which $G(\frac{S}{N},x)$ is the distribution describing the waveform statistics of one and $H(\frac{S}{N},x)$ is the distribution describing the waveform statistics of the other. In comparing the performance characteristics of two processors with different output statistics, ROC (Receiver Operating Characteristic) curves are useful.

A ROC* curve[5] is a graph of probability of detection, D, plotted as a function of clutter rate, the processor input signal-to-noise ratio being held constant. The ROC* for a typical processor has the form shown in the sketch at the left where clutter rate is plotted on a logarithmic scale. When the ROC* of a second process is sketched (dotted) for the same input signal-to-noise ratio on the same plot, the superior processor is clearly the one whose ROC* appears on the left, corresponding to lower clutter rates for a given detection probability.



6.4.2 A Detection Performance Curve

An easier method of comparison is to plot the input signal-to-noise ratio required of a processor, so that its detection probability is 0.50, as a function of clutter rate. Two nomographs for constructing plots of this type for the replica correlator and the envelope detector have been designed by H. R. Courts. They appear in Figs. 6-21 and 6-22. The curves at the left, labeled false alarm rate (another term for clutter rate in this discussion), give the threshold, expressed in dB above the average processor output noise power relative to the processor output mean, which must be used to maintain the clutter rate at a desired value. The parameter on the curves in the correlator nomograph is pulse bandwidth; the parameter on the curves for the envelope detector nomograph is pulse length. These curves describe the statistical behavior at the output of the processors for noise alone when two-dimensional Gaussian noise

[5]The ROC curve is usually "Detection probability plotted against probability of occurrence of a clutter point. This type of ROC curve is inadequate for processor comparison. In this write-up ROC* will be used to denote a plot of detection probability as a function of clutter rate; it is sometimes called a modified ROC.

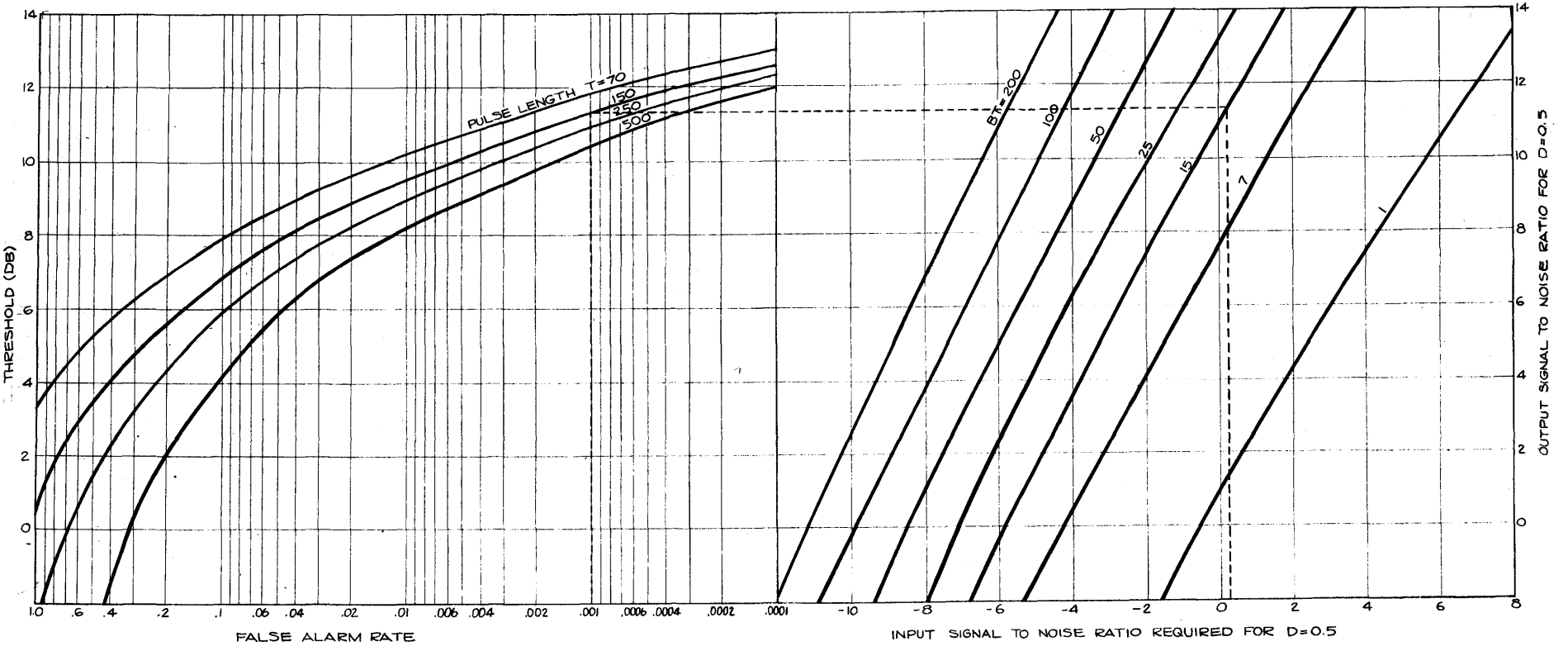


FIG. 6-21 ENVELOPE DETECTOR DETECTION NOMOGRAPH AT 0.5 DETECTION PROBABILITY

TRACOR, INC. 3 DWG 768-98
AUSTIN, TEXAS 4-29-63 BMB-ES

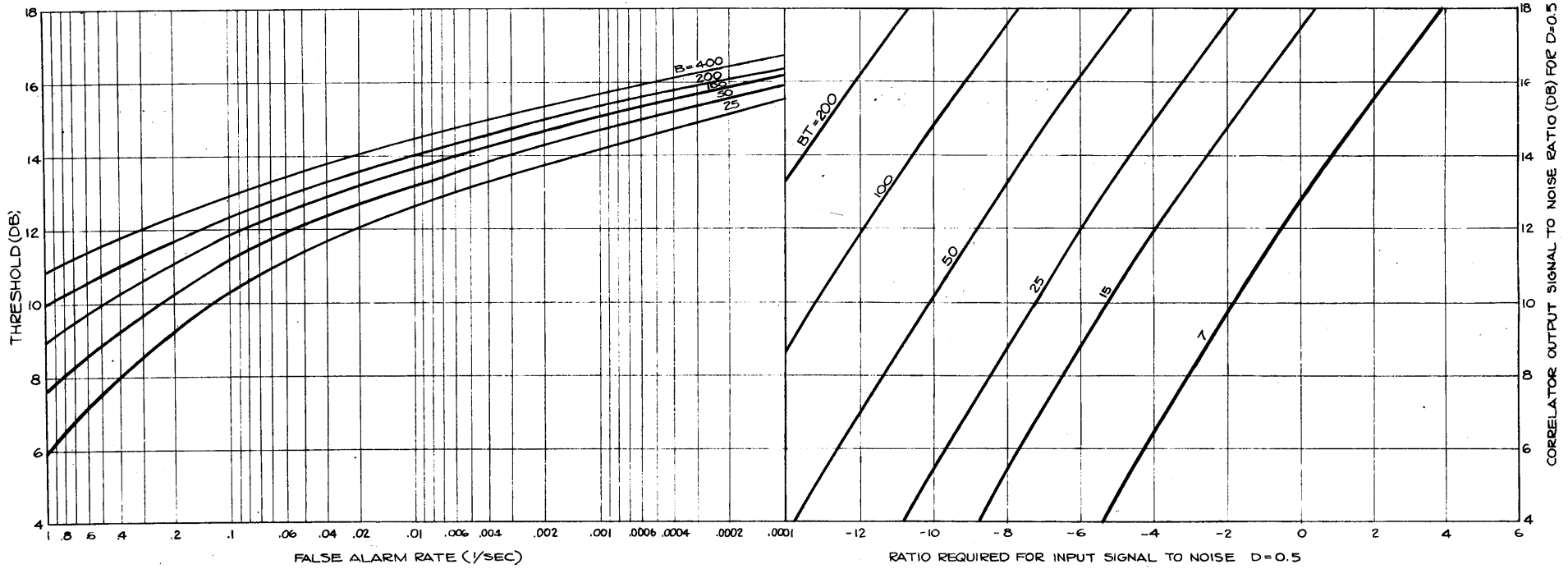


FIG.6-22-CORRELATOR DETECTION NOMOGRAPH AT 0.5 DETECTION PROBABILITY

appears at the processor inputs. The envelope detector curves are based upon Gaussian statistics; the correlator curves are based upon Rayleigh statistics.

The curves on the right of the nomographs are coupled to those on the left on the vertical axis, the threshold axis on the left in dB being the same scale as the output signal-to-noise ratio on the right. The parameter on the curves on the right is BT, the input bandwidth times the pulse length. These curves provide the connection between output signal-to-noise ratio and the input signal-to-noise ratio for 0.5 detection probability.

An example of the use of the curves is shown dotted on Fig. 6-21 for the envelope detector. From the desired clutter rate a vertical line is drawn to the curve corresponding to the pulse length employed, (bandwidth for the correlator) then a horizontal line is drawn from this point to the BT product for the signal employed, then a second vertical line is drawn from this point to the input signal-to-noise ratio axis. The clutter rate and corresponding input signal-to-noise ratio required for 0.5 detection probability are found. Figure 6-23 shows typical detection performance curves prepared for a linear correlator and an envelope detector with 500 ms, 100 c/s FM slides. The correlator can operate in this manner with 2-3 dB lower signal-to-noise ratio than the envelope detector.

If multipath splitting occurs all of the input signal will not appear in a single resolvable interval of the correlator output. For example if three equal spikes appear at the correlator output the required input signal-to-noise ratio must be interpreted as that available for producing one of the resolved output spikes. The total input signal-to-noise ratio must therefore be about 3 dB = 10 log 2 above the value read from the curves to have 0.5 detection probability. Under these conditions the correlator and the envelope detector require almost the same input signal-to-noise ratio for 0.5 detection probability.

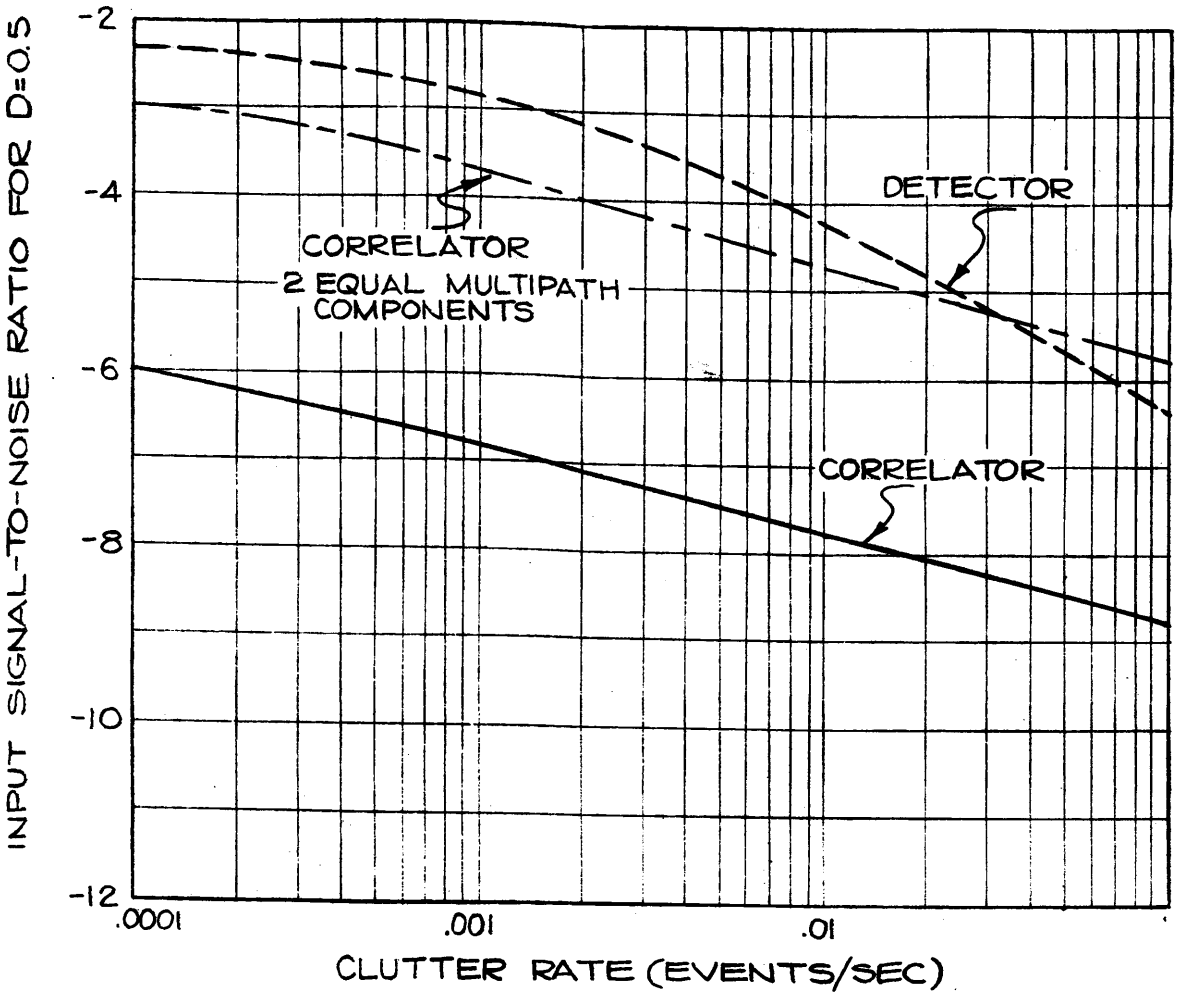


FIG.6-23 - DETECTION PERFORMANCE CURVES: REPLICA CORRELATOR-ENVELOPE DETECTOR

TRACOR, INC. AUSTIN, TEXAS | DWG. 771-22 | 5-24-65 BMB-ES

In this example, comparisons have been made in terms of the "input" signal-to-noise ratios. At a clutter rate of ~~0.01~~ 0.1 events per second the correlator, in the absence of multipath, requires an input signal-to-noise ratio of -8.8 dB, and the detector-averager requires -6.4 dB, for 50% probability of detection. The correlator is able to detect a signal 2.4 dB lower in signal-to-noise ratio than the detector.

No mention of the processing gain has been made in this discussion. This example provides an excellent illustration of the way in which processing gain can be misleading. To see this, the "output" signal-to-noise ratios are also obtained from Figs. 6-21 and 6-22. They are tabulated in dB, as follows:

	Clutter Rate	Input S/N for D=0.5	Output S/N for D=0.5	Processing Gain
Correlator	~~0.01~~ 0.1	-8.8	12	20.8
Detector	~~0.01~~ 0.1	-6.4	4	10.4
Difference	0	-2.4	8	10.4

The processing gain, the difference between the output and input signal-to-noise ratios, is also tabulated for the two processors.

The thresholds of the processors are set to provide identical clutter rates and probabilities of detection. Referred to the input signal-to-noise ratio, the correlator is more effective by 2.4 dB. To provide this performance, the correlator has, and must have, 8 more dB output signal-to-noise ratio than the detector. It should be noted that this equivalent performance is obtained with a difference of 10.4 dB between the correlator and detector processing gains.

In terms of the output signal-to-noise ratios or the processing gains, the correlator seems several dB better in performance than the detector. Since this 8 dB advantage at the processor output provides only 2.4 dB advantage referred to the processor input, it is clear that most of the advantage in gain

is "used" in the correlator to compensate for its relatively higher output bandwidth and Rayleigh waveform statistics.

6.4.3 A ROC* Example

One additional example of detector comparisons will be given: the comparison of the detection performances of the replica correlator and the human listener. In this example a predicted ROC* will be prepared for each processor. The basis of the comparison is the statistical behavior of steady signals in two-dimensional Gaussian noise.[6] The time function representing signal plus noise is

$$u(t) = A \sin 2\pi ft + x(t) \sin 2\pi ft + y(t) \cos \pi ft, \quad (6\text{-}96)$$

in which A is the amplitude of the signal and x(t) and y(t) are two independent Gaussian random variables. The probability that x(t) (or y(t)) is less than a threshold ζ is

$$P(x < \zeta) = \frac{1}{\sqrt{2\pi N}} \int_0^{\zeta} e^{-q^2/2N} \, dq \, . \quad (6\text{-}97)$$

In Eq. (6-97) N is the average noise power. It is easily shown that Eq. (6-96) can be written

$$u(t) = r(t) \sin[2\pi ft + \varphi(t)] \, , \quad (6\text{-}98)$$

for which the probability $G(\frac{S}{N},\zeta)$ that r(t), the amplitude of the signal plus noise, exceeds a threshold ζ is

$$G(\tfrac{S}{N},\zeta) = e^{-\frac{S}{N}} \int_{\zeta}^{\infty} e^{-\frac{q^2}{2N}} I_o(\sqrt{2}\sqrt{\tfrac{S}{N}}\tfrac{q}{\sqrt{N}}) \, qdq \, . \quad (6\text{-}99)$$

---

[6] See for example J. J. Lawson and G. E. Uhlenbeck, Threshold Signals p. 51, et seq. and p. 154 et. seq. McGraw Hill Book Co. (1950).

In Eq. (6-99) $\frac{S}{N}$ is the ratio of the average signal power $(\frac{A^2}{2})$ to the average noise power N, $I_o(a)$ is the modified Bessel function $J_o(ia)$ of order zero. When the average signal power is zero, Eq. (6-99) reduces to

$$G(0,\zeta) = \frac{1}{N}\int_{\zeta}^{\infty} e^{-\frac{q^2}{2N}}\, qdq = e^{-\frac{\zeta^2}{2N}} \qquad (6\text{-}100)$$

$G'(0,\zeta)$, the derivative of $G(0,\zeta)$ with respect to ζ, is the Rayleigh density function. This distribution gives the probability that the amplitude of noise alone will exceed the threshold ζ. The distribution of Eq. (6-99) gives the probability that the amplitude of the signal plus noise will exceed the threshold ζ.

The distribution $G(\frac{S}{N},\zeta)$ has been tabulated by Marcum.[7] Graphical representation of the distribution appears in Figs. 6-24 and 6-25. These are identical plots except for the scales on the axes. The thresholds are expressed in units of 2N and the probabilities plotted give the probability that the signal plus noise amplitude exceeds ζ. The probability axis is labeled "Detection Probability." The plot labeled $\frac{S}{N} = -\infty$ dB is the plot for noise alone.

The ROC* curve for the replica correlator will be prepared first. The signal form will be assumed to be a 150 ms, 100 c/s FM sweep. It will be assumed that a clutter rate F of 0.001 per second (about one clutter point every 16 minutes) is a satisfactory operating condition. The output bandwidth of the replica correlator is B = 100 c/s, thus a statistically independent sample of noise is received on the average every $\frac{1}{B}$ = .01 seconds. There are to be on the average about 0.001 clutter points per second so that the probability of clutter P in a single independent noise sample must satisfy the equation

---

[7] J. I. Marcum, "Tables of Q-Functions" Project Rand Research Memorandum M-339, 1 January 1950, The Rand Corporation, 1700 Main St., Santa Monica California. ASTIA Document Number AD116551.

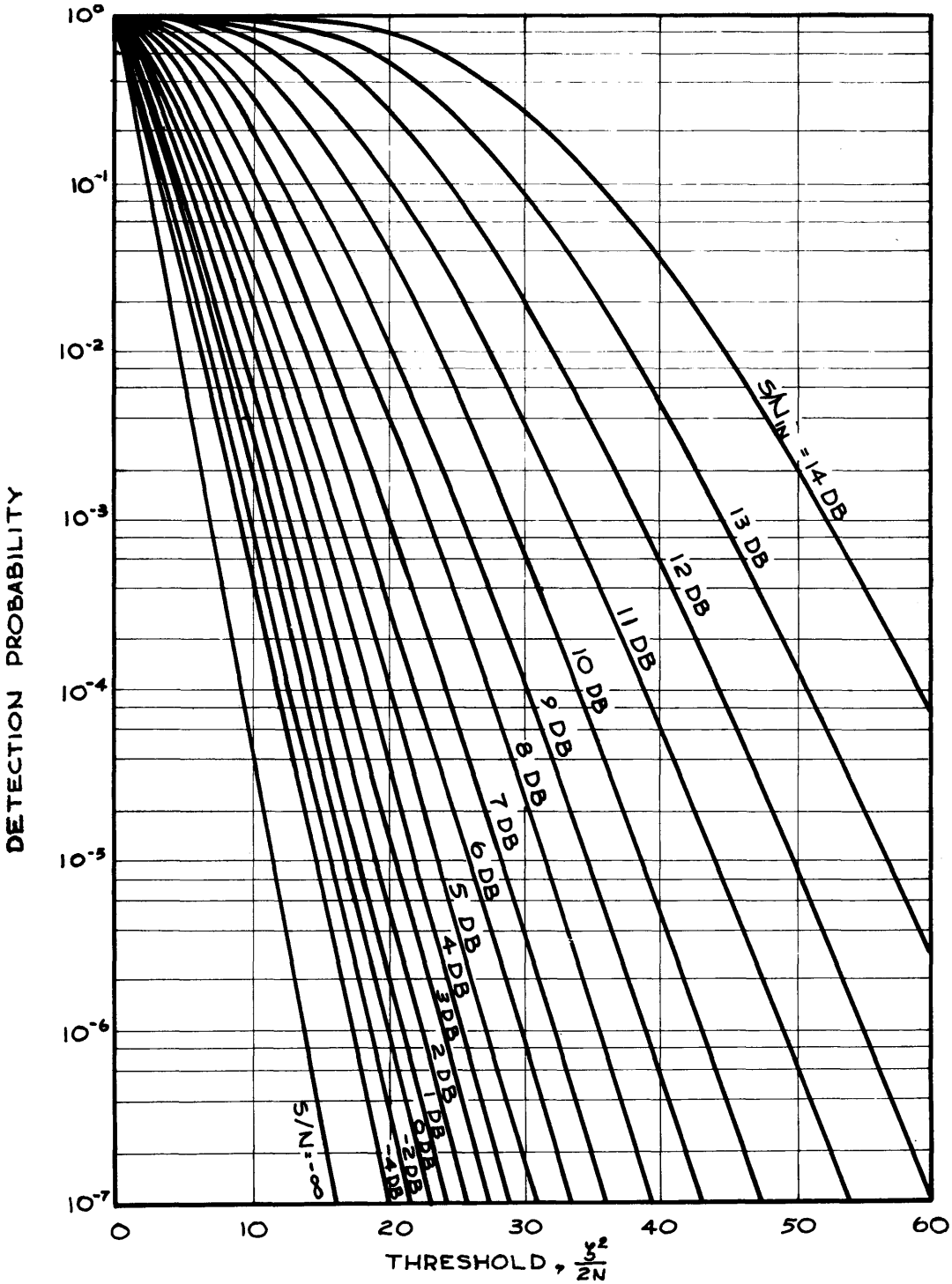


FIG. 6-24 - PROBABILITY THAT SIGNAL PLUS NOISE POWER EXCEEDS v^2

TRACOR, INC. | DWG. 768-123
AUSTIN, TEXAS 10/7/65 BMB/LBH

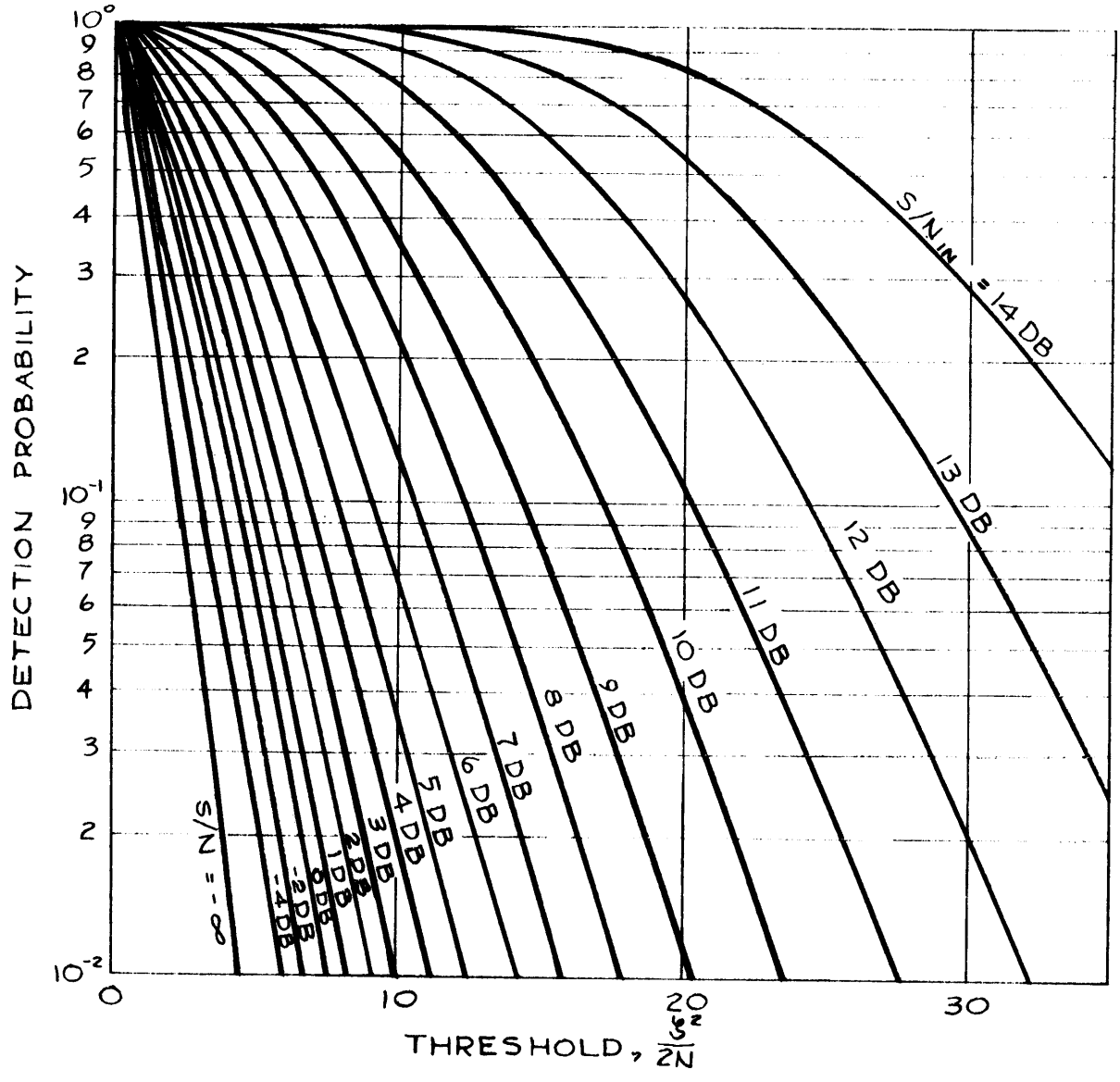


FIG. 6-25-PROBABILITY THAT SIGNAL PLUS NOISE POWER EXCEEDS V^2

TRACOR, INC. | DWG. 768-122
AUSTIN, TEXAS 10/7/65 BMB/LBH

$$F = BP \ , \qquad (6\text{-}101)$$

the rate of receiving independent samples multiplied by the probability of producing a clutter point with a single sample. It follows that

$$P = \frac{F}{B} = \frac{.001}{100} = 10^{-5} \qquad (6\text{-}102)$$

Figure 6-24 is entered on the "Detection Probability" axis with $P = 10^{-5}$. The threshold value given by the $\frac{S}{N} = -\infty$ curve is $\frac{\zeta^2}{2N} = 12$. This threshold setting is required if the clutter rate F on noise alone is to be .001/second on the average.

The detection probability as a function of signal-to-noise ratio is read at the intersections of $\zeta^2/2N = 12$ and the curves corresponding to different signal-to-noise ratios. The scale of Fig. 6-25 is more convenient for carrying out this step. Figure 6-26 shows detection probability plotted as a function of signal-to-noise ratio obtained for the conditions assumed in this example. The signal-to-noise ratios obtained from Figs. 6-24 and 6-25 are the signal-to-noise ratios at the input to the "threshold detector" shown in the block diagram of a signal processing system in Section 6.4.1. The signal-to-noise ratios are then those at the output of the "Processor" - in this case the signal-to-noise ratios at the output of the replica correlator. Two other plots, those for F = .01/second and F = .0001/second are also shown in Fig. 6-26.

In order to compare the performance of this processor with another it is necessary to convert the output signal-to-noise ratios to signal-to-noise ratios at the input of the processor. In the correlator

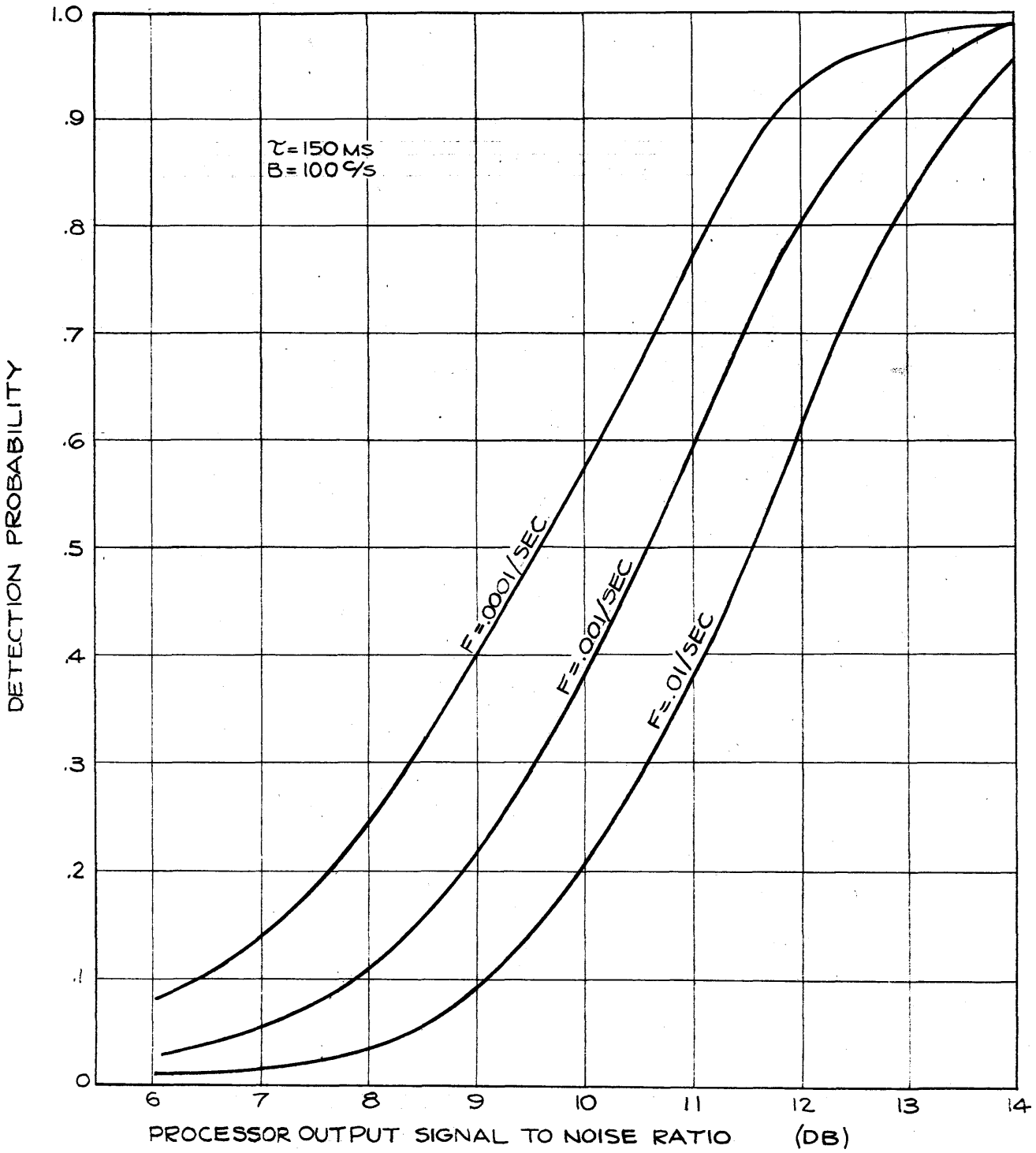


FIG. 6-26 - CORRELATOR DETECTION PERFORMANCE

TRACOR, INC. AUSTIN, TEXAS | DWG. 768-85 | 4-28-65 BMB-ES

$$\left(\frac{S}{N}\right)_{out} = 10 \log BT + \left(\frac{S}{N}\right)_{in} , \qquad (6\text{-}103)$$

when average signal-to-average noise ratios are written. For the pulse form which is being used in this example

$$10 \log BT = 10 \log (100)(.15) = 11.8 \text{ dB}. \qquad (6\text{-}104)$$

Figure 6-26 can now be redrawn in terms of correlator input signal-to-noise ratios by reducing each point on the horizontal axis by 11.8 dB. The new plot is given by the solid plot in Fig. 6-27.

The other processor, the human listener, has been found equivalent to a 40 c/s bandpass filter[8] which can track the FM sweep. He is therefore like one of the processors in the block diagram in which a 40 c/s filter is in the processor box along with a heterodyner to convert the sweep to a CW. His processing gain for the 100 c/s FM sweep is therefore the ratio

$$G = 10 \log \frac{B}{40} = 10 \log 2.5 = 4 \text{ dB} . \qquad (6\text{-}105)$$

The output bandwidth of the listener is 40 c/s so that he receives an independent sample of noise every $\frac{1}{40} = .025$ second. If he is to maintain his clutter rate at F = .001 per second, his probability P of false detection on a single independent noise sample must be held at

$$P = \frac{.001}{40} = .000025 . \qquad (6\text{-}106)$$

[8] These are the approximate parameters which describe the equivalent behavior of the human auditory system when signals are in the neighborhood of 800 c/s.

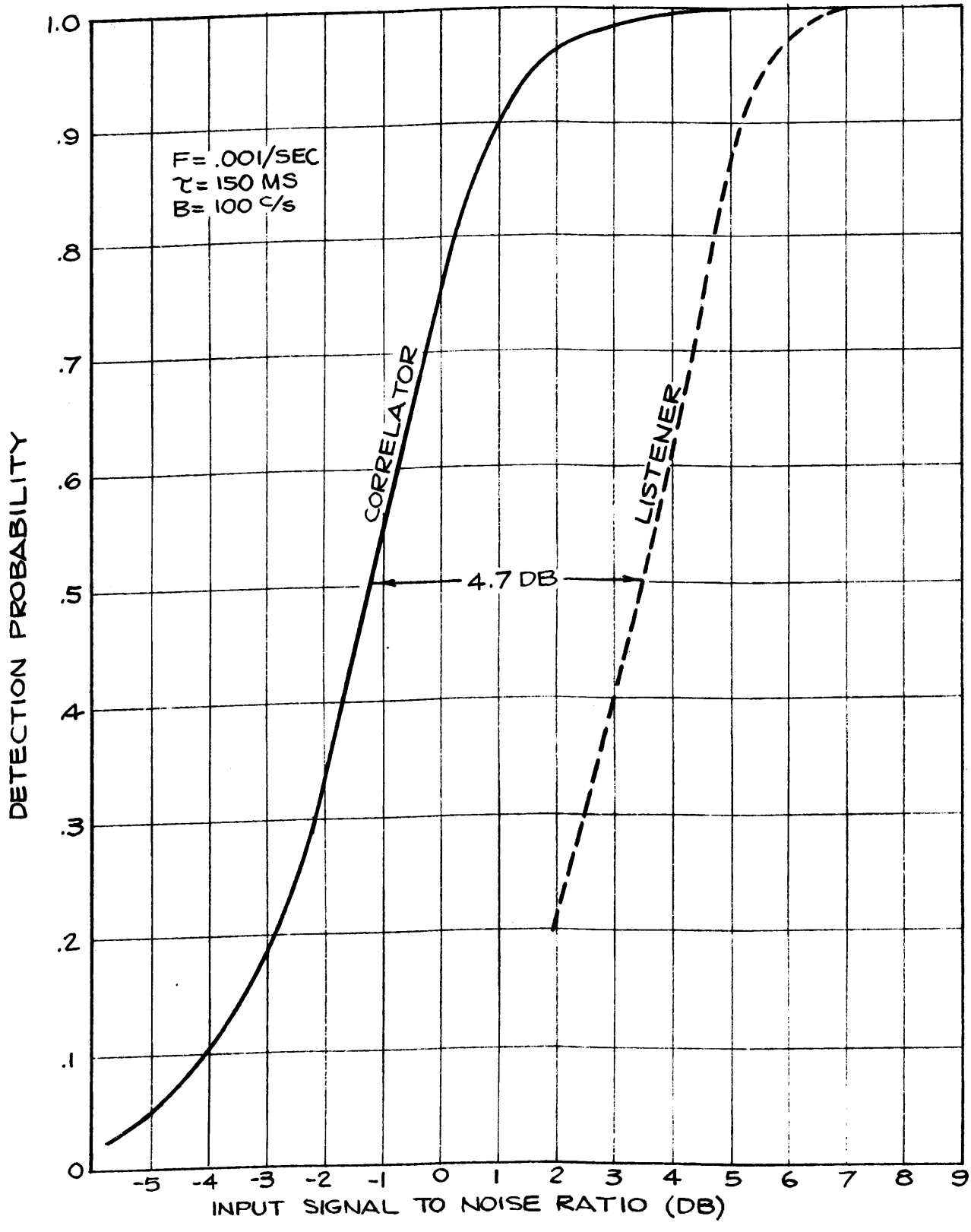


Fig.6-27-DETECTION PERFORMANCE: HUMAN LISTENER - REPLICA CORRELATOR

TRACOR, INC. AUSTIN, TEXAS | DWG.768-87 | 4-28-65 BMB-ES

According to Fig. 6-24, the threshold required to give this probability of clutter is $\zeta^2/2N = 11$. The probability of detection $G_1(\frac{S}{N},\zeta)$ as a function of "processor" output signal-to-noise ratio is read from the intersection of the curves for different S/N ratio with the threshold $\zeta^2/2N = 11.8$. This detection probability D_1 is given the subscript 1 to denote that this is the detection probability of a signal in a single independent noise sample.

The listener example differs from the correlator example in two ways: (1) The listener has a lower processing gain. (2) The output signal duration in the case of the listener exceeds the time between independent noise samples. The difference in processing gains tends to make the listener performance poorer than the correlator performance. The extended signal duration provides more opportunities to detect because if the output signal length is T and B_o is the output bandwidth, the listener will have $T/(1/B_o) = TB_o = k$ opportunities to detect the signal, that is, the signal will appear in TB_o different noise samples. His probability D_k of detecting in at least one of the k samples is

$$D_k = 1 - (1-D_1)^k \qquad (6\text{-}107)$$

In Eq. (6-107) $1-D_1$ is the probability of not detecting in a single noise sample, $(1 - D_1)^k$ is the probability of not detecting in k samples, and its complement is the probability of detecting in at least one of the k samples. A plot of D_k against the "output" signal-to-noise ratio is shown in Fig. 6-28.

When the output signal-to-noise ratios of Fig. 6-28 are transformed with Eq. (6-105) the resulting plot as a function of input signal-to-noise ratio is the dotted plot of Fig. 6-27. The replica correlator detection performance is superior to the listener performance by about 4.7 dB, i.e., the replica correlator can detect 150 ms signals 4.7 dB lower in signal-to-noise ratio than the listener at a specified clutter rate.

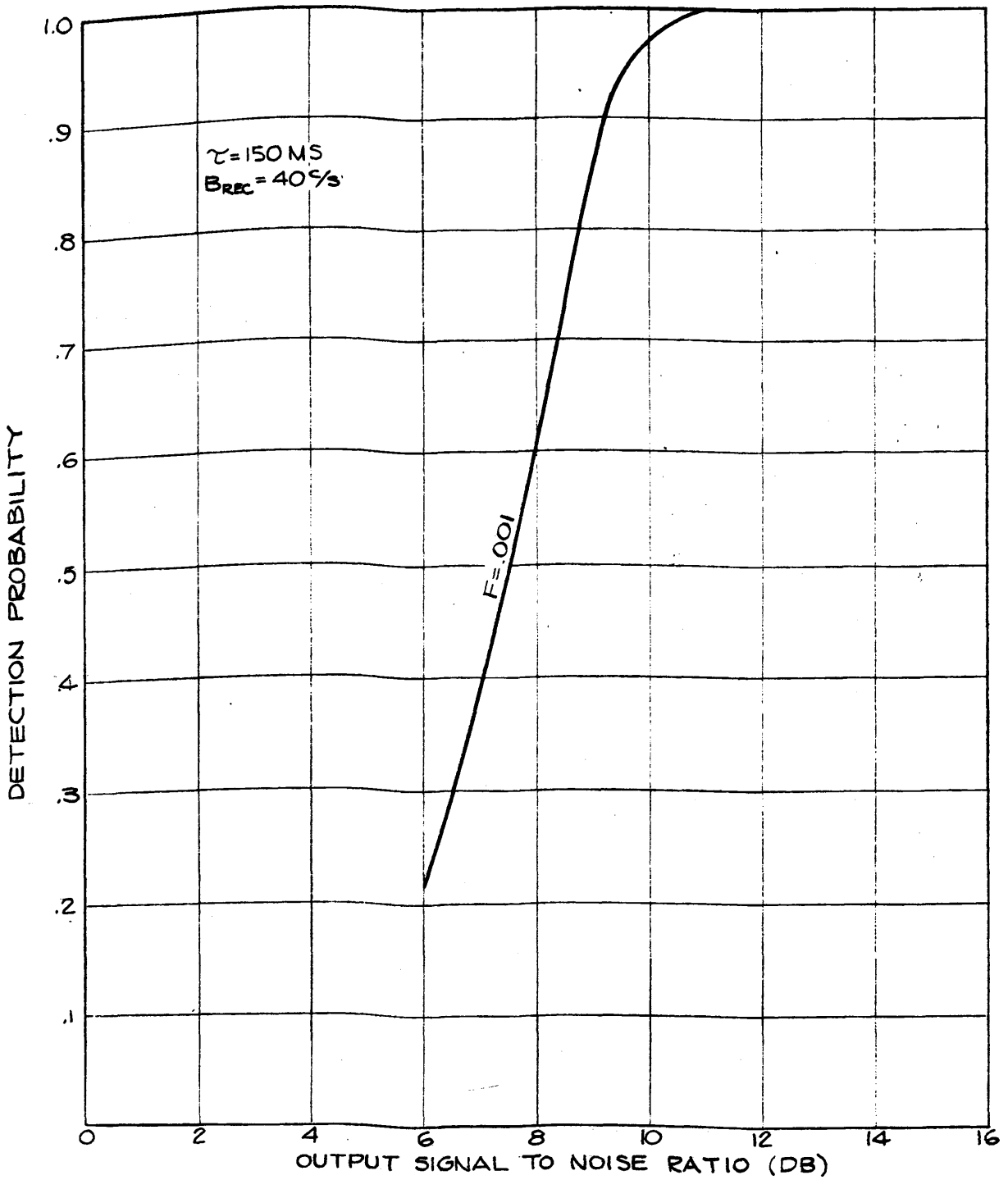


FIG. 6-28 DETECTION PERFORMANCE: HUMAN LISTENER

TRACOR, INC. AUSTIN, TEXAS | DWG.768-86 4-28-65 BMB-ES

6.4.4 Other Processors

The processors which have been discussed are

1. Replica Correlator
2. Detector Averager
3. Human Listener

Several other processors which are variations of the replica correlator and the detector averager are of importance. One is the so-called "spatial" correlator. In this processor the output of the receiving array is split into a right-half output and a left-half output. The two outputs are correlated in real time. The product of the two signals is averaged over the expected signal duration if active signals are being processed and over a time in which statistical stationarity can be expected when processing passive signals. Success of the technique depends upon zero correlation between the two outputs for noise alone and unity correlation in the absence of noise when the array is precisely aimed at a distant localized source. The output correlation coefficient is proportional to

$$\text{arc Sin } \left(\frac{S}{S+N}\right) \quad .$$

Except for the 3 dB degradation in signal-to-noise ratio which occurs in splitting the transducer this processor would be superior to the detector averager and would be equivalent in performance to the replica correlator had it one noise-free channel.

The ping-to-ping integration is a form of post detection averaging whose statistical description can be based upon convolutions of the distributions plotted in Figs. 6-24 and 6-25. These convolutions have not been carried out so that a

specific example cannot be given. There are difficulties in implementing ping-to-ping integration except by means of a multiple trace A-scope presentation in which the human eye performs the integration. This type of processor increases E in Eq. (6-83) by adding energy from succeeding pings.

There are also clipped versions of the correlator processors. Clipped signal processors provide nearly optimum processing with simplification of such problems as automatic level control and normalization before applying thresholding techniques.

7. CONCLUDING REMARKS

7.1 PASSIVE SONAR

The operation of passive sonar can be described in terms of well known principles of space-time signal processing. The attainable processing gain, and hence the available output signal-to-noise ratio, depends on the nature of the acoustic sources, the nature of the propagation of the energy from these sources to the receiver, and the nature of the noise background.

The space processing or beamforming is normally the first element of a passive sonar. The attainable processing gain or directivity index depends on the size of the receiving array, the shape of the signal wavefront, and the characteristics of the noise field at the receiving array. The classical approach to beamforming is to assume a plane signal wavefront and an isotropic incoherent noise field. Under these assumptions a theoretically optimum coherent beamformer can be designed within the limits imposed by the size and shape of the receiving array. The beamforming process can be carried out using either analog or digital techniques. The predicted output signal-to-noise ratio of the beamformer may not be realized if the above assumptions are not realized. Improved performance must then come from measuring and compensating for deviations from the assumed signal and noise characteristics. Signal wavefront curvature can be measured and compensated using adaptive signal processing. The effectiveness of adaptive processing or improvement in beamformer performance will depend entirely on how often significant wavefront curvature occurs. In a similar fashion the noise field characteristics can be measured and compensated for in the vertical and horizontal planes, using adaptive processing, with improvement depending on the deviation from the assumed isotropic noise field. That is, a coherent noise source can be detected and the beamforming can be adapted to be optimum in the presence of this noise. A coherent

noise field is entirely analogous to the multiple target problem.

A passive sonar must cover a relatively wide band if a large fraction of the target signal is to be admitted to the time-processing networks. Since sonar beam characteristics are related to both the acoustic frequency and the array dimensions, conventional beamforming techniques are inadequate for wideband signals. In a system using preformed beams to cover an angular sector, if the beams at high frequency just fill the sector, the beams at low frequency will be badly overlapped. Conversely, if the beams at low frequency just fill the sector, there will be serious gaps in coverage, as well as harmful multi-lobing at the high frequencies.

These variations in directivity index cause a greater discrimination against the noise field at high frequencies than at low. The noise spectrum at the array output thus has a more negative slope than exists in the water, and this creates problems within the signal processing system.

The difficulties described above can be eliminated or significantly reduced by providing an array and beamforming network which yield a frequency-independent beamwidth. Techniques to accomplish this have been adapted from antenna theory [1-3].

Multi-lobing can be overcome, to a large extent, by tapered spacing of the elements. The taper is chosen so that element-to-element differences in arrival time, off the main lobe, cannot simultaneously be an integral number of wavelengths for a large number of hydrophones. To overcome the variations of

1. D.D. King, R.F. Packard, and R.C. Thomas, "Unequally-Spaced, Broad-Band Antenna Arrays," IRE Trans on Antennas and Propagation, AP-8, 380 (July 1960).
2. M.G. Anderson, "Linear Arrays with Variable Interelement Spacing," IRE Trans AP-10, 137 (March 1962).
3. A. Ishimaru, "Theory of Unequally-Spaced Arrays," IRE Trans AP-10, 691 (November 1962).

beamwidth and directivity index with frequency, a frequency-dependent beamformer is also required, to maintain an approximately constant aperture-wavelength ratio over the frequency band of interest.

Because of the consistency in beamwidth and lobe structure, the slope of the noise spectrum at the array output will approximate that of the noise in the water. In addition to its other advantages, an array incorporating these principles requires fewer elements than a conventional array.

Following the beamforming several different kinds of time processing may be used. The passive sonar has become increasingly more sophisticated, progressing from "incoherent" processors to one or more forms of "coherent" processors.

The incoherent processor is a rectifier followed by an averager, i.e. a power detector. If a signal appears with noise on a particular bearing, while only noise appears on the other bearings, the output signal-to-noise ratio (peak on the bearing containing signal compared to the average noise fluctuation on the other bearings) differs from the input signal-to-noise ratio by the "processing gain." For small signals the processing gain is proportional to the available integration time and the input signal-to-noise ratio. For large signals the processing gain is proportional to the integration time. This leads to the unfortunate situation that the incoherent processing gain decreases when it is needed most, i.e. for small signals.

The processing gain can be increased by increasing the integration time. To accomplish this long integration a transition from scanning to fixed or preformed beams has been made in practice. The maximum available integration time is determined by how long a target can be expected to stay on one bearing. This time is of course dependent on the beam pattern, but a value of 1 second is typical. To obtain further improvement by increasing the integration time one must sum all possible combinations of bearing

corresponding to all possible target paths. Some progress has been made in performing this long term integration electronically, but it has been and is still being done using an operator scanning a display. GE serial DIMUS has a display which carries an hour's worth of data for scanning by an operator. Improved techniques for performing this multiple summing process electronically, and an evaluation of operator performance, are still of interest.

An improvement is obtained by employing a coherent processor. One form of coherent processor uses the cross correlation between the two outputs of a split transducer array. This technique provides a processing gain which is independent of the input signal-to-noise ratio except in a transition region between large and small signals. In theory, 3 dB is lost in the space processor by splitting the array into two halves, but this 3 dB is recovered in increased gain of the time processor. The net result is to obtain the same over-all gain as the incoherent detector for large input signals without the severe decrease in gain for small input signal-to-noise ratio. This technique provides more processing gain for small signals than the incoherent passive system provided the acoustic field originating at the target is highly coherent at the two halves of the array and only if the background acoustic noise field is not. The role of adaptive time processing and long integration time is the same as described above.

Still further improvement in the passive time processor can be achieved if a priori knowledge of some of the characteristics of the passive signal can be specified. In the limiting case where the detailed structure of the expected signal could be known, a 3dB improvement would be obtained for large signals and a 6 dB improvement would be obtained for small signals. This could be done by using the cross correlation between the beamformer output and a noise free signal, thus avoiding the 3dB loss from splitting the receiving array. The amount of improvement which

can be obtained in the operating environment depends primarily on the extent to which one can predict the structure of the target acoustic field. Adaptive processing can be employed to make this prediction.

To this point we have been considering the continuous acoustic field radiated from the target. Submarines are known to radiate transient acoustic energy as a result of necessary functions aboard the target. As permanent components of the target passive signature, they will be a factor in describing the passive detection capability. However, it should be noted that passive transient signal detection is really an opportunistic technique.

A possibly important approach to predicting the target signature for passive sonar (or even the noise or reverberation behavior in the active case) is second or higher order statistics. This means primarily achieving an understanding of joint probabilities, i.e. given the noise or the signal now and in the past, what can be learned from this to predict the new or next value of the noise or the signal plus noise. This is a new area and can be pursued efficiently with present day computer capabilities.

7.2 ACTIVE SONAR

As with passive sonar, improved active sonar performance over and above that attainable with a "single omnidirectional hydrophone with a visual or audio display" can be achieved by carrying out spatial signal processing, i.e. beamforming, and temporal signal processing, here coherent processing by matched filtering which can be followed by incoherent processing or ping-to-ping integration. Beamforming, done by either analog or digital techniques, both for transmit and receive, is analogous to that for the passive case. Adaptive beamforming, to compensate for non-plane signal wavefronts and for non-isotropic noise can probably be limited to the receive array. Ultimate spatial

processing gain is limited by the number of independent samples of signal and noise that can be obtained, and, except for the possibility of booms attached to ships or submarines, the ultimate limit is represented by an array which covers the hull of a ship or a submarine (planar array). Note that independent samples of noise can be obtained by sampling the field at separations of about a wavelength or perhaps a little less[4]. For matched filter (replica correlation) processing, it is well known that the best attainable output signal-to-noise ratio is given by

$$E/\eta \ ,$$

where E is the energy in the target echo and η is the noise power per unit bandwidth. (This is the best we know how to do right now theoretically.) Thus we want to make E as large as possible, which means that we want the largest transmitted power and longest pulse duration possible. However, two things happen which limit this procedure. First, as we increase the transmitter power level, we become cavitation limited, so we need to stay just below this level. Second, as our pulse duration increases, say to about 70 ms for a 140 dB source level, we become reverberation limited. This means that η is now determined by the reverberation level, and further increases in transmitted signal duration increase E and η at the same rate. The best place to operate is the point at which we are just barely reverberation limited.

We have a gimmick though. Note that when we are reverberation limited, we can reduce η but keep E constant for a given pulse duration if we increase the bandwidth of the transmitted pulse (remember that η is the power per cycle/second). Thus, in principle, i.e. if we had the available power plant on board, we could make the transmitted pulse very long and for each duration, find a bandwidth which just barely made us reverberation limited. Here is a place where discussions with transducer people are necessary.

---

[4] J.J. Faran and R. Hills, Jr., "Correlators for Signal Reception," Tech. Memo. No. 27, Acoustics Research Lab., Harvard University (Sept. 15, 1952).

This sounds real good, but there is a catch. As we increase the bandwidth, recall that we increase the range resolution. This means that our output from the matched filter gets shorter and shorter, i.e. our long broadband transmitted pulse should get stacked into a smaller and smaller time interval at the processor output. But, as luck would have it, the medium, its boundaries, and the echo structure produced by the target are such that the echo energy arrives with a distribution in time. Thus, instead of the echo continuing to stack up as we increase the bandwidth we soon reach a point where we get a lot of smaller echoes, resolved and side-by-side. For detection and using bottom bounce we have evidence to indicate that the limiting bandwidth is at about 200 c/s. This problem needs to be investigated for all anticipated modes of operation.

The energy splitting produced by the medium, its boundaries, and the target for high resolution pulses can be compensated for if we learn how to design a transmitted pulse, or if we learn how to design correlator references, which will again "match" the filter and give us a single output. This is an essentially unexplored field. We can also sum the multiple outputs incoherently, i.e. use a detector averager, but the net worth of this is unknown. We have evidence which shows that, for bottom bounce data, although the signal peak can be increased in this way, the noise peaks are also increased since they have the same structure as the signal, both produced by the layering of the bottom. This points out the great need for developing the necessary tables and curves which can be used for predicting false alarm rates and false alarm probabilities corresponding to all of the commonly used or proposed processing techniques, so that ROC curves can be conveniently obtained to correspond to signal excesses or signal-to-noise ratios calculated from the sonar equations.

Along with pushing each of the above efforts to reasonable limits, another course of action remains to increase target detection probability or decrease false alarm rate. This is ping-to-ping integration. As in the passive case, one can, in principle, do a lot of automating here, but the best ping-to-ping integrator is apparently still the operator. Many approaches can be taken to the automation process. At this time the main approaches used are those which reduce the amount of work the operator has to do. An example is the preselector or computer aided post-detection processing. The value of this and the extent to which it should be pursued needs study, both in terms of interaction with the system and also with the operator.

Classification and identification can follow along the same lines as the detection problem. Here, however, wider bandwidth pulses than for detection may be useful, provided that the medium and medium boundary effects can be removed. This needs further study.

Besides the above own-ship considerations, there is the problem of best use of our knowledge and technology considered in terms of multi-ship operations for a variety of operational requirements. Reverberation models need to be constructed not only for the own-ship situation, but also for the "bi-static" case, so that significant analyses can be performed on the multi-ship sonar performance interaction problems.

☆U. S. GOVERNMENT PRINTING OFFICE : 1966 O - 211-710

NAVSHIPS 0967-129-3010

ERRATA SHEET

Entered

PAGE	LOCATION	ACTION
12	5 lines from bottom	Change "most" to "lost"
16	2nd equation	Change to "$(\frac{\text{dyne}}{\text{cm}^2})/(\frac{\text{cm}}{\text{sec}})=(\frac{\text{g cm}}{\text{sec}^2\text{cm}^2})/(\frac{\text{cm}}{\text{sec}})=\frac{\text{g}}{\text{cm}^2\text{sec}}$"
26	Introduction to Eq. (2-12)	Should read " ...the ratio of the measured or actual intensity (I_m) to the predicted intensity (I_p) at some point..."
27	Lines 2-6	Delete redundant paragraph
36	3rd line from bottom	Change M_R to M_N
52	Illustration	Right hand Z should be Z_L
52	7th line from bottom	Change "techology" to "technology"
60	Line after Eq. (3-17)	Change $\gg$ to $\ll$
141	Line 13	Change "given" to "give"
145	Line 4	Change "silimar" to "similar"
154	Eq. (6-40)	Change to $\|F(f)\|^2 = \left[\frac{A^2T^2}{4}\right]\left[\ldots\right.$
154	6th line below Eq. (6-40)	Change "90%" to "80%"
154	Bottom	Change "continguous" to "contiguous"
178	Eq. (6-76)	In 2nd line of equation, insert bar over entire expression in brackets. In second term in brackets, change subscript $\Delta/2$ to Δ/τ, where appearing. Under double summation, add "$i \neq k$".
179	Eq. (6-82)	In denominator of left-hand side, change subscript "uu" to "uw". Add a bar over "$u^2(t)$" in numerator of right-hand side.

ERRATA SHEET (cont.)

PAGE	LOCATION	ACTION
186	Line 13	Add "a" between "filters"and "significant"
208	Line 2, and Table	Clutter rates should read 0.1 for both Correlator and Detector.
209	Eq. (6-97)	Change $\sqrt{2\pi n}$ to $\sqrt{2\pi n}$